

国际信息工程先进技术译丛

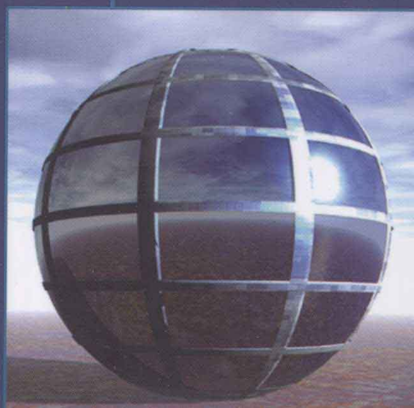
现代通信原理

（原书第2版）

Kommunikationstechnik

(2. Auflage)

（德） Martin Meyer 编著
李玉宏 等译



机械工业出版社
CHINA MACHINE PRESS



国际信息工程先进技术译丛

现代通信原理

（原书第2版）

（德）Martin Meyer　编著
李玉宏　等译

机械工业出版社

本书系统地介绍了现代通信原理，包括通信系统的基本概念、原理和结构，数字传输的基本原理和技术，调制和编码的相关概念、原理、技术方法，以及无线传输和有线传输等各种传输系统的原理、技术和传输媒介等，最后还对各种通信网的基本原理及所涉及到的技术进行了详细的解释。

本书可作为高等院校通信与信息系统专业的本科生教材，也可作为通信行业的工程技术人员（电信运营商和设备制造商等）的参考书。

Martin Meyer: Kommunikationstechnik (2. Auflage)
Original ISBN: 3-528-13865-3

图书在版编目（CIP）数据

现代通信原理/（德）迈耶（Meyer，M.）编著；李玉宏等译．—北京：机械工业出版社，2010.12
（国际信息工程先进技术译丛）
Kommunikationstechnik
ISBN 978-7-111-32505-5

Ⅰ.①现… Ⅱ.①迈…②李… Ⅲ.①通信理论 Ⅳ ①TN911

中国版本图书馆 CIP 数据核字（2010）第 222787 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）
策划编辑：张俊红 责任编辑：朱 林 责任校对：陈立辉
封面设计：马精明 责任印制：杨 曦
保定市中画美凯印刷有限公司印刷
2011 年 1 月第 1 版 · 第 1 次印刷
169mm × 239mm · 23.75 印张 · 530 千字
0001—2500 册
标准书号：ISBN 978-7-111-32505-5
定价：88.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务	网络服务
社服务中心：（010）88361066	门户网：http://www.cmpbook.com
销 售 一 部：（010）68326294	教材网：http://www.cmpedu.com
销 售 二 部：（010）88379649	
读者服务部：（010）68993821	封面无防伪标均为盗版

译 者 序

目前，各种网络应用迅猛发展，随之而来对通信系统的要求是能够提供高带宽的、实时的、可靠的双向数据传输。为了实现这样的通信系统，要使用数字传输、调制、编码等多种技术手段，采用同轴电缆、光纤、无线和卫星传输等多种传输技术，同时针对不同的需求构成局域网、城域网、广域网等不同的网络形式。

本书是一本关于通信原理的导论性质的书籍，它的内容基本涵盖了实现上述通信系统所用到的技术。本书在解释通信技术的各个基本概念的同时，还注重强调这些概念在整个通信系统中的作用，并强调了由不同概念组成的通信系统的各个环节之间的相互关系。同时，本书还对通信技术中所用到的理论知识进行了适当的介绍，详略得当。本书既给出了所用到的理论的主要内容和性能，又没有特别详细的、让初学者感到头疼的理论推导，同时还为今后的深入学习打下了基础。本书还包含大量的图形、图片和表格，材料新颖丰富，内容翔实，兼备系统性、知识性、可读性、指导性。

本书共分为6章。第1章对通信原理做了全面的、总体上的介绍，引出了通信系统的总体结构。第2~4章分别对通信系统中所用到的具体技术，包括数字传输、调制和编码等进行了详细的描述。第5章根据实际的通信系统对这些技术的共同影响进行了分析，并解释了具体的传输媒介。最后本书在第6章中对基于通信技术的数据网进行了概要的介绍。

本书由北京邮电大学李玉宏老师翻译。在翻译的过程中，北京邮电大学的硕士研究生宋肖肖、侯云静和庄秀斌审校了部分初稿，并更正了初稿中的很多书写错误，在此向她们表示衷心的感谢。此外，邴佳楠、邓伟、刘宝文、李一山、高维国、李丽平、吴井义、李韬、姜树嘉、蓝锦娜、唐超超、段家琛、高桂芬、李守信、夏亦婷、潘淑文、郑秀丽、郭小莉、赵义、徐琳、袁韬、张涛、李齐、修艳芳、田梦华、李国华、时占等也参与了资料整理、部分翻译及相关支持工作，在此一并向他们表示真诚的谢意。

由于译者水平有限，加上翻译时间十分仓促，因而本书翻译中的错漏之处在所难免，恳请各位专家和读者不吝指出。

译　者

第2版前言

只有当接收端具备了通信的条件，并且发送端和接收端相互合作时，通信才可能进行。人与人之间的信息交流同机器之间的信息交换是一样的。跨越空间和时间距离的通信需要技术手段的辅助，这就是本书所要讨论的主要内容。

通信技术是令人着迷但又难管理的领域——每一个子领域的技术均进步飞快，以至于连专家几乎都跟不上前进的步伐。人们以前使用电话网进行语音通信，而现在传输数据是主要的。电信、计算机科学与娱乐电子技术的相互融合，其不可阻挡的趋势是向全球的、交互式的、移动多媒体方向发展。对技术人员来说，这意味着双向的、宽带的、实时的，基于光纤、无线和卫星传输的广域网上的数据通信。

本书的通信技术导论包括涉及到的这个领域的大部分内容，偏重于介绍数字概念。有些部分的描述涉及一些理论，因为抽象的理论比实际中使用的方法更通用、更持久。但空洞的概念必须要在实践中落地，因此本书经常会给出一些实际实现中的数值。当然这些数值要小心使用，因为它们变化非常快。尽管如此，这些数值对目前仍有很重要的参考价值。

本书适用于学生和从事实际工作的、特别是电子技术和信息技术领域内的工程技术人员，同时也适用于对通信工程感兴趣的自然科学工作者。对理解书中的内容来说，跟其他的理工科基础课一样，数学方面的基础是必不可少的。同时也需要一些信号处理方面的知识，因此有关通信技术的书籍通常会包含关于“信号和系统”一章的内容（傅里叶变换、拉普拉斯变换、z变换，频率响应、脉冲响应、零极图等）。因为信号处理有足够多的资料，所以我在这里省略掉了这一章。如果读者需要这些内容，可以参考参考文献【Mey02】或【Mey00】。对于实现和执行书中介绍的概念，还需要计算机科学、数字技术、电子学和电子技术方面的知识。

这里所涉及的系统绝大多数还处于结构图阶段，它们只用于表明相互之间的关系。例如包含低通滤波器这个模块，但这个滤波器如何实现（有源还是无源，模拟还是数字，使用纯硬件还是使用信号处理器等）我们并不太关心——信号处理、数字技术和电子线路等专业会专门负责对它们进行讨论。

本书包括通信技术所涉及到的所有重要的概念。通常情况下，一个概念的要旨只有在一定的环境中才能领略到。因此，完全理解每个部分概念并不是很重要的，重要的是要搞清各个模块相互之间的关系以及对整个系统的影响。因此本书的整个内容是按照需求逐渐升高的方式进行组织的。

第1章是对整个领域的一个导言。第2~4章分别对数字传输、调制和编码的

相关概念进行详细的描述。第5章根据实际的通信系统对它们的共同影响进行分析，并仔细研究具体的传输媒介。最后本书在第6章中对数据网技术进行了概要的介绍。从结构上来说，就像一栋大楼一样，第1章是整个大楼的基础，第2~4章涉及的是像暖气、供水等这样的设施，第5章从整体上对这幢楼进行描述，第6章以居住区的描述作为结束。

本书的重点在于它的广度和相互之间的关系，许多地方都要参考其他相应的段落，为此需要作下记录。因此我推荐不要很慢地、很仔细地阅读此书，而是快速地阅读两遍。这样能够更好地利用相互之间的参考，并弄清总体上的关系。

读者学习完本书之后会对通信技术有个基本的和实时的了解，但更多的是个全面手而不是专家，接下来还需对每个专门知识进行持续不断的学习。为了使进一步的自学成为可能，还需要增加一定的基础知识，这应该能从读者自身得到保证。应该有意识地注意用英文表达的一些专业概念，这样读者在阅读大量的英文文献时会觉得容易些。

关于统计通信理论，本书只涉及一些基础的内容，并且只以直觉的方式给出。这些理论与传统通信的关系就像量子物理与传统物理之间的关系一样。没有坚实的基础知识，则学习这些特殊理论是没有意义的。但对后面的工作来说，统计通信理论是一个很有价值的领域，本书给出了关于这些内容以及对后面特殊领域的入门知识。

本书第2版对第1版的一些书写错误进行了纠正，对部分解释的表达进行了精炼，并对一些内容进行了更新。

在此感谢与 Vieweg 出版社同事的愉快合作。

本书包含我在 Aargau 技术大学（瑞士）的一些教学工作。

马丁·迈耶

目　录

第1章 基本原理

1.1 信息传输导论

1.1.1 通信技术的内容及范围

不论是在有生命的还是无生命的自然界，各种组织及子系统之间都相互在不断地进行交换。比较明显的是物质和能量的交换，但不太明显的还有信息的交换。根据现代的观点，这3种交换上相互之间是平等的。

通信技术涉及的是远距离的信息交换，为此它需要多种技术系统。

具体来说，信息交换涉及到下面内容的传输：字符、文字、数据、话音（Voice）、音频（Audio）、静止图像和活动图像（Video）。

信息的例子有娱乐、监测、登记、命令等。信息的交换和处理靠信号的帮助来进行，这里信号是对信息的物理上的描述。

信息是知识的内容，信号是对知识的描述。

因此信息传输只能出现在可观察到的事件是可感知的情况下，这个事件有任意的物理大小（即信号），可感知、能反映接收设施的敏感性。

由于信息与信号是相互关联的，因此信息的固有内涵到很晚才被发现。香农在1948年创立的信息论中对信息的描述和量化进行了讨论。

从本质上来说，任意的物理信号对信息进行描述都是适合的。例如人的体温是一个信号，它能反映所涉及到的人的健康状况。温度可用光学的方法来表示（温度计上有颜色的点），但也可用机械的（水银柱）、声学的（声音的高低或大小）、电的（电压或电流的大小或频率）方法来描述。物理描述方法可以在不篡改信息内容的前提下进行改变，因此人们选择对信息传输、处理、存储等最适合的信号类型——通常是电信号，因为它们：

1）有较高的传播速度，能远距离传输；

2）有较高的带宽，因此可以包含更多的信息；

3）不一定非得与介质联系在一起，从而可以实现无线传输；

4）很容易操纵，如利用物美价廉的立体声设备、电话机、计算机等。

人们早期将电子信号传输与信息传输等同起来，因此电信技术是电子工程的一个传统领域。由于信息论已经发展成一个独立的学科，与此同时电信技术也成为数学家和计算学家喜欢的领域。对电子工程领域来说，电信技术也成为面向数学的一门学科。于是下面的问题就有严格的不同：

1）对信号必须使用什么样的算法？

2）怎样才能实现一个执行上面算法的系统？

为此有一个例子，信号传输可通过下面的方式来进行：空间的距离→电信，时间的距离→存储。

对理论数学来说这两种情况根本没有什么区别。实际上卫星通信采用与 Compact Disc（CD）几乎相同的编码方法。但在实际实现中会遇到很多的不同，因为电信号相比于存储来说更利于传输。相应地在 CD 中存储信息跟在模拟唱片上存储信息一样都是通过刻录表面来完成的，即借助机械信号。能源技术有着同样的困惑：有一个有效的系统来传输能源，但它却使用机械的办法来存储能源，例如存储在抽水储能站水库中。同样缺少合适的存储设备的电动车还不能取得突破。

电信号与光信号的竞争越来越激烈：使用同轴电缆传输 15000 路电话话音每公里需要 600kg 的铜，而使用光纤的话 100g 的玻璃就够了！同时光信号传输是一种已经建立起来的技术。采用光系统来进行信号处理的革命性进展也是值得期待的。虽然光同样是电磁波，但光学是物理学的一个分支。这里用什么和怎样来说明一下两者分离的好处：

1）已经成熟的理论不会很快地被新技术搅得天翻地覆，它的老化是慢慢地进行的。例如对于理论人员来说，数字信号是用同轴电缆来传输还是用光纤来传输是无所谓的。

2）理论可以推动应用的发展，因此能为技术打好基础。例如 CD 就是这样，它的理论在 1950 年就已经为人所知了，但设备却是在 30 年后才出现的。

3）一种理论能从其他的领域吸取知识或向其他领域输出内容。因此自动控制技术与电信技术有很多共性，使用同样的数学描述方法是很常见的。

电信技术所涉及到的领域很宽，已变成不同领域专家的嬉戏场。信息工程师理论上出现问题时可以得到数学、信息论、系统论等方面专家的支持，而在实践中出现问题时可以得到电子、物理、计算机等方面的工作者的帮助。只有当信息工程满足下面的条件时，上面的做法才有可能成功。

1）了解其他学科的基本特点；

2）拥有互通的能力。

正是强调人与人之间的交流（这里主要指语言方面的能力）的第二点经常会出现问题。

图 1-1 描述了不同专业领域之间的关系。用粗线框起来的部分是传统电子工程专业的学习内容，而用细线框起来的内容由于下述原因还没建立起学习研究的专业：

1）与其他学科合并（遥测、远程信息处理）。

2）总称（传输、电信技术、计算机科学），这里显然是通过其他的方法来描述的。注意图 1-1 并不是构建细微差别的基础，而是显示它们之间的关系！

3）集成到其他学科中（网络理论、场论）。

4）专业知识（信息论、交换）。

值得强调的是信息系统需要所有专业（包括能源技术）的公共合作才能完成，而

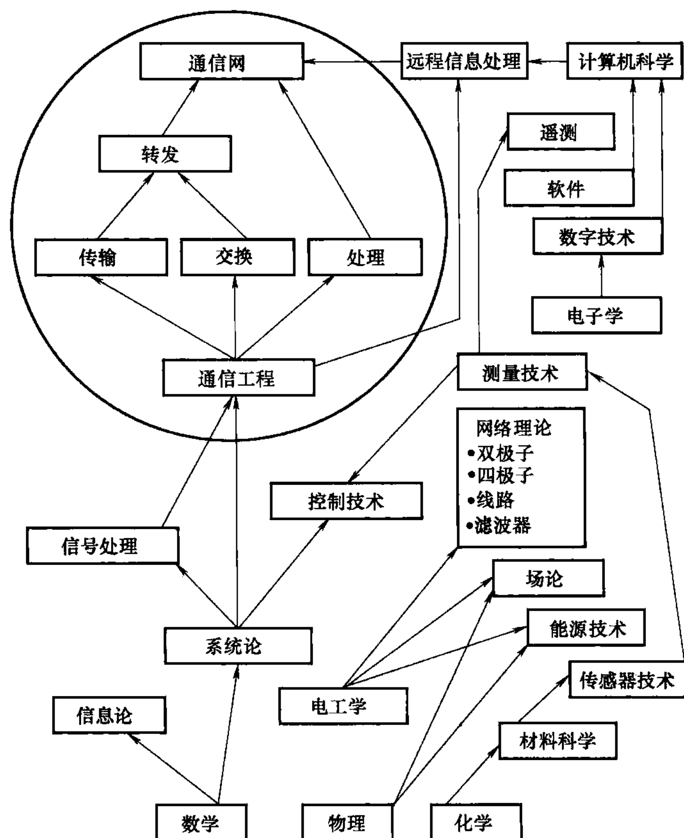


图 1-1 电信技术与其他知识领域的关系

电信技术像自动控制技术和计算机技术一样又在向其他的领域渗透。

特别要强调的是电信技术（Telecommunication）与计算机技术相结合产生了信息技术。两者巨大的区别涉及到典型的理性观点[Tra93]：

1）年龄的区别约为 100 年；

2）两者结合之前电信技术经历了较慢的技术上的发展，而计算机技术年轻且相应地发展较快；

3）电信技术在国家的垄断下成长，而计算机领域的公司都是私人企业；

4）结合前电信技术处于模拟世界中，而计算机技术对模拟世界一无所知。

两者的合并曾经有并且现在也有足够好的原因：

1）目前两种技术都需要同样的背景技术，即微电子和软件；

2）每种技术为了自己的目的都需要另一种技术，即计算机技术需要电信技术的帮助实现计算机联网，而电信技术中对交换机和其他设施的监测和控制也需要计算机技术的支持。

以前人们认为电信技术主要就是信号的传输。这个分支在电子管时代是电子电路技术的最重要的应用。同样交换技术实际上更多的是精密机械技术而非电子技术（如中继器）。此时信息处理还根本是不可能的。大的电信公司相应地从组织上也划分为传输技术部门和交换技术部门。

随着集成电路的出现交换技术发生了根本的变化。今天传输、交换和处理使用一种相同的技术，那就是高度集成的数字技术（微电子技术）与占很大一部分比例（现在已经约为50%）并且这个比例还在不断增长的软件相结合的技术。除此之外，人们看待问题的方式也发生了改变：电信网不再是由与传输设备相连的交换机组成，而可以看成是许多分布的交换机。根据这一观点，许多电信公司都进行了重组，将传输和交换又合成了一个部门。

电信技术（或者更宽一点来说是信息技术）已经变成了一种很关键的技术，人们经常会听到信息社会这样的说法。因此，虽然不管是否适合读者，下面的各项内容均可以当成知识来看待：

1）电信技术不是“小且优美”的技术，而是一门范围很广的技术。它的发展计划已经变得极其昂贵（几亿欧元的数量不是不常见的）。这样的后果是（排除小公司）只有国际化的大联合体才有生存的能力。因此，标准化变得很重要，但很遗憾它的意义经常被低估。并且虽然标准化是技术领域的工作，但它实现起来却更像交易和外交工作。

2）电信技术有很重要的经济意义。因此只做应用者而不做发展者是很危险的。

3）电信技术对人类的日常生活有很重要的影响。因此也会产生社会方面和文化方面的影响。为此可以引用1980年瑞士PTT通信基本原理的10条方针“PTT已经意识到，不是所有的都是技术上可行的、经济上可以承受的并且是社会需要的，电信领域的发展要从整体上进行判断。”

对上面这些内容的解答不能故意放慢发展的步伐，只能是支持更多的主观独立性。多数情况下主动参与是必要的，但这对过程的参与者要求是很高的（图1-1也应该进行扩展，以便将外语、社会学、经济学、政治学、企业管理学、伦理学等内容包含进去）。这里将通信问题首先变得明朗化是正确的，因为很遗憾许多工程师宁愿放弃他非常喜欢的专业领域也不愿意从一个门外汉变成一个明白人。这需要有许多人，他们同时具备一般性的能力而非特殊的能力、沟通的能力和自我表现的能力。

作为一般了解：1994年9月底瑞典电信拿出将近5千万欧元，目的是使所有的员工（由于早期欧洲市场的开放只有30000人）都能够获得瑞典语、英语、数学以及视不同工作领域物理或企业管理等专业的大学入学考试的资格。这些人接受教育是必需的。引用继续教育部领导的话：“在电信领域简单工作的时代肯定是过去了”。

1.1.2 通信技术的基本方法

本节将在后面各个章节的详细介绍之前，先勾画出一个整体结构。这里将介绍所有的基本概念。许多基础知识在这里认为是已知的，或只是很简单、很肤浅地进行介绍，详细的介绍将在后面各章进行。本节的目的只是给出方向性的概况。

> 传输系统的任务是将信息从信源传送到信宿。后者与传输信道或简称为信道相连。

图 1-2 给出了这个原理。以前是人作为信源和信宿（例如电话通信），现在通常由机器来承担信源和信宿（例如进行电子财务处理的数据传输）。

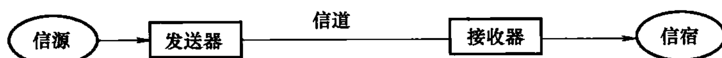


图 1-2 传输系统的简单结构

其实人类之间的通信也符合图 1-2。人与人之间的通信与技术之间的通信一样，发送者和接收者之间必需高度的配合。它们最大的区别是技术之间的通信（发送者和接收者是机器）传输的是信息，而人与人之间的通信有一个最大的目标，就是相互之间能够互相理解。当然这里所说的是技术通信。电信工程传输的是信号，它自己并不关心信号的内容（信息）。就跟信使传递信件和包裹一样，他们并不关心里面的内容。电信技术提供一种业务，即连接。使用者（如果他付费的话！）使用这种业务，即进行通信。

使用者对传输系统进行评价的标准是：源的保真性（“音的质量”）、可用性、可靠性、范围、速度、完整性（私密性、窃听）和价格。

当然所有的评价标准都能同时得到最大的满足是不可能的。很多情况下使用者必须说明他想要什么样的质量（根据实际情况和主观情况来评价）。然后根据这些要求以及相关信道的特性设计出一个最优的传输系统。经济方面使用大炮来射麻雀当然无论如何都是没有意义的。

同样的原理也适用于信道，它的目的不只是信息传输。物质的传输（如货物的传输、包裹的邮寄）同样有不同的可能性，用户选择一种对他的目的是最优的方案。例如在邮寄包裹时有不同的速度需求（正常邮寄、快递、专人速递）和/或不同的安全性需求（正常、挂号、专人投递）需要得到满足。没有人会简单地将一切东西都用专人来速递。

人们希望传输系统做的就是不出差错地传输信息。但实际上人们做的仍然是传输信号。所以现在就变成信号在传输的路径上绝不允许出错！尽管信号出一点错仍然能正确地传输信息。这就像虽然艺术手写体被很笨拙地誊写，但仍然能看出句子的意思。这个例子也说明，实际的有用信息是通过定义得到的。是否应该将写满字的纸的外观送到接收者还是将文字所表达的内容交给接收者是有区别的。

信号通过一个可变的电压、压力及温度等物理方法描述信息。数学上信号是用函数来表示的。绝大多数情况下使用的是一维信号，即函数有唯一的一个独立变量。在图像处理中函数使用两个独立的位置参数。因为物理形式对理论思考是没有影响的，通常将信号写成没有维度的 $s(t)$。此外，时间是独立的变量。在一个音乐磁带为例中长度 x 是独立的变量，如果音乐磁带使用恒定速率 v 向音频磁头写的话，那么利用

$t=x/v$，与长度无关的磁化强度就转换成与时间有关的电压。

包含信息的信号是随机信号。美国人 Hartley 早在 1927 年就认识到了这些并要求电信工程应该多使用随机信号来代替正弦信号进行工作。很有讽刺意味的是正是这个 Hartley 以发明 Hartley 振荡器（一种正弦信号发生器）而著名。确定信号是可以计算的，它的曲线是可以预知的。如果人们计算出一个信号在晚一些时刻是什么样子，那么这个信号的传输就可以立刻停止，因为可以直接在接收器中产生这个信号。但绝大多数作为传输技术的敌人的噪声类型的干扰信号都是随机信号。这意味着信息论和传输理论与概率计算和统计学有很大的关系。但两个学科给在决策思想方面经过培训的工程师并没有带来多少同情。相反确定信号在简单的测量技术和简单的数学表达方面有很多优点，因此在电信技术中使用得很多。

如果信号载有信息，那么一个或多个信号参数（幅度、相位、频率等）就会随信号发生变化。可以证明：

> 每秒钟传输的信息越多，需要的带宽就越大，并且信号变化得就越快。

上面的两种说法是等价的，这从傅里叶变换的特性中就可以得到。表 1-1 很清晰地表明一幅图的内容多于 1000 个字！

表 1-1 不同信源的典型带宽

信号类型	带宽/kHz	说明
远程打字机	0.025 ~ 0.2	视数据传输速率而定
话音	2 ~ 10	视希望的质量而定
音乐	20	
电视	5000	
辅助信号	0.005 ~ 1000	视类型和任务而定

信源和信宿被时间上和空间上的距离分开，电信技术的所有努力都是围绕战胜这些展开的（见图 1-2）。不是每个信号都能在一个给定的信道中传输的，多数情况下要求信号和信道必须相互适合。例如在一个很吵的机器设备中间口头上的交流是很困难的，但交换记录的纸条却很容易。在一个很黑的房间中人们不能阅读，但理解声音的内容是一点儿问题也没有的。除了信号的特性例如带宽之外，还必须识别信道的特性，它们包括例如下面的值：

1）信道衰减：这个值规定了发送器输出端所必需的最小功率以及没有中间放大器（转发器，再生器）所能跨越的最大距离。

2）干扰：在从信源到信宿的路径上，信号会受到各种类型噪声的干扰（包括试图阻止通信的故意干扰或欺骗）。根据应用的不同，还必须要保证私密性。

3）频率的上、下极限：这些值决定了可用带宽以及由此确定的每秒钟能够传输的信息量。表 1-2 给出了一个例子。

表 1-2　不同传输介质的频率范围（指导值）

介　质	频率范围	应　用
双绞线	0 ~ 100MHz	电话(用户范围),LAN
同轴电缆	0 ~ 1GHz	TV、电话(交换机之间的连接)
光波导	200 ~ 375THz	数据、电话(交换机之间的连接)
无线	10kHz ~ 1THz	无线广播、定向广播、卫星、雷达

C. E. Shannon（香农），信息论的创始者，给出了一个信号的最大信息含量的公式。它是（在 1.1.8 小节给出了详细的分析）

$$I = B \cdot T\log_2\left(1 + \frac{P_S}{P_N}\right) = B \cdot T \cdot D \tag{1-1}$$

式中　I——信息量，单位为 bit；

B——信道的带宽，单位为 Hz；

T——传输时间，单位为 s；

D——动态性，用于表示干扰的敏感性；

P_S——信号功率；

P_N——噪声功率（N = noise）。

式（1-1）隐含了下面重要的内容：

带宽、传输时间及抗干扰性可以相互转换

例如一句话可以说三遍，虽然这持续的时间会长些，但却可以保证传输误差能被纠正。另一个例子是磁带-录制的音乐可以使用双倍的速率来传递。此时音调会增高一倍（带宽也会翻倍），但传输时间会减半。此外在不改变信息量的情况下，可以使用更多的带宽来保证无干扰的传输。这点已被 UKW 无线广播充分地利用：正是基于这个原因这里充分利用了频率调制（FM）。在中波和短波无线广播中 FM 虽然技术上是可行的，但由于在这个频段内，带宽比较紧张，为了进行干扰保护必需使用对频谱资源比较经济的幅度调制（AM）。也可以使用另外一种表达方式：如果要在中波无线广播中使用同样的具有抗干扰功能的调制方法，那么只有少数的几个发送者可以发送。具有相同信息量的数字信号比模拟信号需要更大的带宽，虽然这是个缺点，但这种方法比 FM 具有更大的抗干扰能力。

为了根除广泛流行的错误我们要记住下面的一句话：

> 对成功的传输来说不一定是接收器输入端的信号功率尽可能地大，而是要信号功率与干扰功率的比值足够大！

这个比值（被称为信噪比）就是式（1-1）中的比值 P_S/P_N。

干扰可以理解为一直都存在的噪声（半导体中和电阻中的噪声、无线传输时大气和宇宙的噪声），也包括人工干扰（人为的噪声、干扰、恶意干扰等）。

弱信号可以被放大。但如果这个信号已经被干扰，那么噪声也被跟着放大了，信噪比得不到改善。出于相同的原因，卫星接收天线的电子放大器直接接在天线上而并没有放在较热的管子上。

在日常生活中这种所需最小信噪比的概念在每个饭店中都会经历到：每位讲话者都根据环境的噪声自动地提高说话的声音，主要是为了让对方能够听得清楚。

通过提高发送功率或减小干扰功率都能够改善较差的通信情况。后者可以通过屏蔽、转换到无干扰的信道（无线传输时）或使用可旋转的天线来实现，此时不是提高对端而是将天线特性调节到干扰最小处。这种方法对消除电视的重影也很有效（重影是由于在某些点进行反射而形成的多径传输所引起的有时延的接收信号造成的干扰）。

现在我们可以更仔细地观察一下传输系统，图 1-3 给出了每个传输系统都包含的单元。

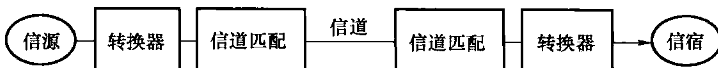


图 1-3 传输系统的基本结构

1）信源的例子包括：人的发声器官、乐器、信号发生器、计算机硬盘等。它分为连续信源（模拟的）和离散信源（数字）。

2）转换器将信源信号转换成另一种物理形式。如将传声器的空气压力振荡转换成电信号。铅笔将想法（=信息）转换成光的形式（纸面上可变的黑色）。

3）通常情况下转换后的信号是不适合直接传输的，它必须经过信道的适配。例如：

① 话音信号所占用的频率范围是 300～3400Hz，这是不能使用天线来输出的。因为无线电波只能在大于 10kHz 时（大于 100kHz 时是比较理想的）才能用便携天线发射。信道适配在发送前将话音频谱转换成更适合传输的频率范围，这可通过调制器来完成。

② 如果要通过电话线来传输计算机信号（数字信源），信号也必须要经过适配。特别是要将信源信号中的直流成分去掉。这也要使用调制器来实现放在 Modem（调制解调器）中。电话系统的线路虽然可以传输直流成分，但为了实现电位分离（阻止均衡电流）而使用了变压器，并且放大器是 AC 耦合（很简单地设置工作点）。因此信道不再能传输直流成分，但这并不限制原始的传输话音的目的。话音信号没有超过 100Hz 的成分。

③ 为了保护信息不受干扰，信道适配是有免疫力的。为此要以带宽和/或传输时间为代价。这个任务还是由调制器来完成，FM 无线广播的例子已经提到过，数字信号情况（例如 CD 的情况）为了实现干扰保护而使用了信道编码。

④ 通常情况下信道具有很高的传输能力，从单个信源来的信息流并不能充分利用它（比较表 1-1 和表 1-2 以及图 1-4）。在这种情况下，将不同信源的信息流汇集在一起并使用同一个信道来传输是很经济的。这被称为多路复用或复用传输。电话网就充分利用了这个原理，利用调制器将不同的话音信号调制到不同的频率范围上传输。人

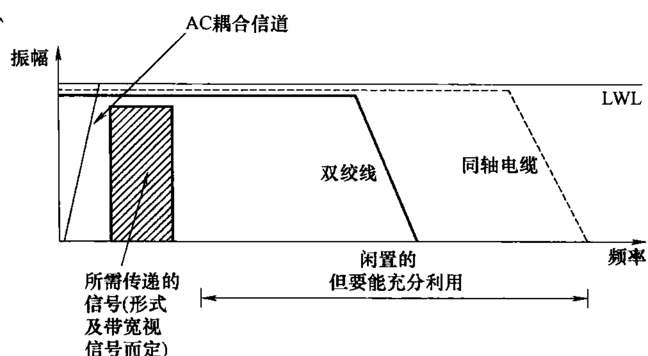


图 1-4 信道带宽和信号带宽

们称之为频分复用（Frequency Division Multiplex，FDM），如图 1-5 所示。邮政也是采用这种方式工作：将目的地相近的信件使用同一列火车来传输，将这种信件区分开的当然不是频率而是信封上的地址。无线广播也是使用同样的原理：它们同时使用（不

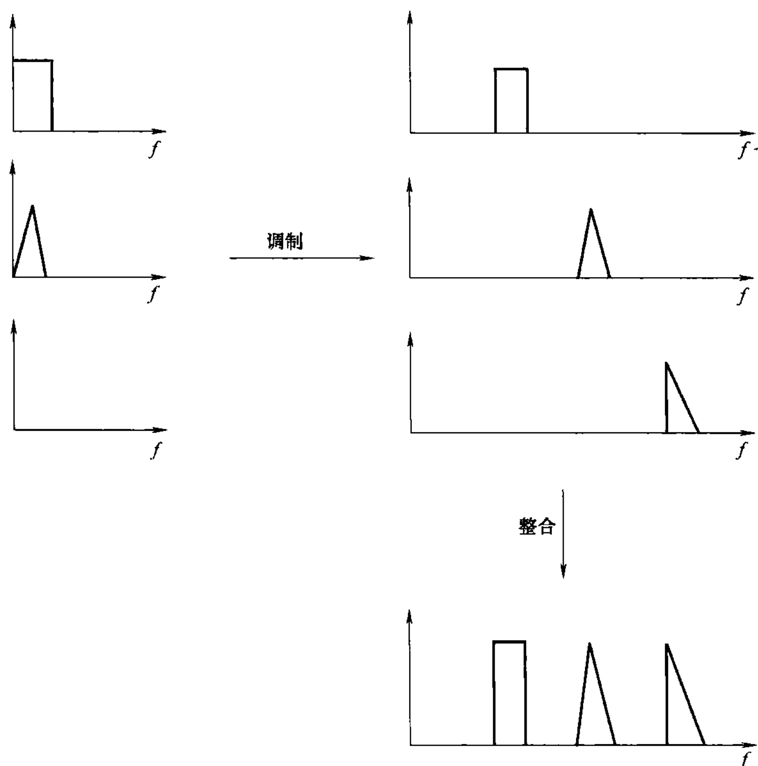


图 1-5 频分复用的原理方法

存在的）以太作为公共的传输信道，接收器通过调节器过滤出希望的发射站。

由于复用电话网需要的电线和电缆（土木工程）大大减少。有了电子技术的帮助（调制器），人们节省的不仅仅是铜，而是许许多多的钱！所有的复用系统都有一个共性，那就是分级的，就像街道网络中，有小区马路、小街道、主干路、高速公路一样。

在接收端，必须将因为信道适配而进行的对信号的变换重新变换回来。话音信号必需重新转换成原始的 300 ~ 3400Hz 的频率范围。计算机信号必须再加上直流成分。这可使用解调器或解码器来完成。在双向连接中绝大多数情况下调制器和解调器都是在一个设备中来完成，简称为 Modem。

通常情况下图 1-3 中的信宿等待的不是电信号，而是例如很好听的音乐。传输系统输出端的转换器负责将信号转换成希望的物理形式，例如音箱、数字唱片、监视器、打印机等就是这样的转换器。

信息论给出了传输系统的物理极限，且目前几乎已经达到了这个极限。可以夸张点说，如果不怕复杂的话，现在任意距离、具有任意安全性的信息传输从技术的角度上来说都是可以实现的。

一般来说，信道是能得到最优利用的。因此在信道适配过程中增加点复杂性是无所谓的，因为电子电路相比没被充分利用的信道来说还是便宜的。最优的信道适配要求掌握有关信道特性的准确知识。第 1.1.7、5.5 和 5.6 小节将对传输信道进行了详细的介绍。

由于传输信道是昂贵的，人们一直在努力寻求只需传输必要内容的方法。一种极端情况是电报。这里关键词是信源编码，它是一种会给将来带来有很大经济意义的技术。信源编码意味着减少冗余和减少不相关。不相关是对接收者来说不产生任何价值的信息，因此接收者根本不希望接收它。例如在音频传输中，传输高于 20kHz 的音频信号是毫无意义的，它最多能用来干扰蝙蝠！在电话系统中甚至会有意识地将话音信号的频率限制在 300 ~ 3400Hz 这个频率范围内。这样话音信号在音色上会有些改变。但经过很多的试验人们已经发现，这样的限制对可理解性上一点也没有影响，说话者仍然能够被很好地识别出来。

冗余意味着是多余的，通常是指一种或者相同的信息存在多份。例如这样的表达“周四，1998 年 4 月 2 日”就是冗余的。1998 年 4 月 2 日就足够了，因为从任何的日历中都能查到这天是周四。信源编码对图像传输是最重要的，电视信号的 5MHz 带宽必须预留，以保证覆盖最大的需求。但绝大多数时间图像都是冗余的。例如当新闻播音员播报文字时，图像中的绝大部分都是不变的。因此只传输嘴部周围的部分图像，图像的其他部分在接收端进行重组就足够了。下面的方法也可以被用到：不是每次都传输完整的图像信息，而只是在第一幅图像中传输完整的信息，后面只传输两幅图像之差即可。这种方法会遇到误差的传递问题。显然，在数据传输中要使用信源编码。此时有趣的一个问题就是图像和话音之间的同步问题：即信源编码需要计算，因此需要时间。话音信号和图像信号的特性不同，因此这两种信号使用不同的压缩方法。尽管如此，提供给观众的必须是经过同步的信号。在模拟电视系统中只是很初级地在颜色传输中使用了信源编码。使用信源编码也能节省传输费用（传输时间和/或传输带宽），

也能够被很经济地利用存储器。

接下来的一个重要概念是信道编码，它意味着通过故意增加冗余来进行干扰保护，当然这所涉及的并不是在信源编码中所消除的那部分冗余。多数情况下增加双份的信息后要能使接收者能够识别，甚至可能纠正错误。但附加冗余后需要较高的传输能力，会增加传输时间或需要较大的带宽。

例如如果一段文字有一个打印错误，那么一般情况下读者可以毫不费力地将这个错误改正过来。这是因为语言是冗余的。相反彩票数字列表就不是冗余的，错误最多是能够被发现（如果遇到两个相同的数字或一个数字太大），但无论如何改正不了。一种最简单的信道编码方法就是将所有的信息都重复发送，在接收器中按照少数服从多数的方式进行判决。但信道编码技术不断发展，通常会采用很巧妙的方式工作。

在对话中根据情况的不同，音节的理解率最多只有80%。在音节理解率是70%的情况下，因为在说话的过程中会存有冗余，所以对话还是可以接受的。当理解率是50%时，还能勉强维持对话。当理解程度变得更差时，经常会需要反复询问。这种“什么?”的反问也可在技术系统中进行实现：信道编码器不是增加了很大的冗余，而是增加了很少的冗余（例如一个奇偶校验位）。如果仅是识别出误码而不能对它进行纠正的话，那么相应的部分还需再传输一次。当然这只有在通信是按照固定的被称为协议的程式进行时才可行，并且还需要有反向信道（但这个反向信道的传输能力可以很小）。这种方法被称为ARQ（Automatic Repeat Request）。

最后还得要说一下加密。这也是一种编码，但它的目的是保密。信息量既没有增大，也没有减少，只是将它变成了另外一种不能理解的方式。

总结：在转换器和信道之间信息信号还要经过下面的阶段：

1）信源编码：通过减少不相关和减少冗余数据量会减小；

2）加密：为了保密而进行的编码；

3）信道编码：为识别和纠正错误而增加冗余；

4）调制：物理信道适配（例如频率范围）、干扰保护和复用。

在接收端将按照相反的顺序执行所有过程的逆过程。

在实际传输系统中当然不需要执行上述的所有过程。图1-6给出的是一个完整的结构图。通常情况下由于信道衰减调制器还需要接一个功率放大器，但以理论的角度上来说这并没有多大意思。

信源编码、信道编码及加密主要是针对数字信号传输中使用的方法。在某种程度上它们的应用理论还比较年轻，这些方法无论如何是不能使用传统模拟技术来实现的。

这里还需要强调：许多“老化的”系统还要被使用很长一段时间。传统方法的相关知识还需要学习，新概念还需要不断地与因不兼容而导致的经济上的障碍做斗争。

许多现象已经表明电信技术所使用的不同方法已经达到一个高度发展的状态。人们不禁要问未来还能出现什么。研究工作将主要在下面的领域展开：

1）信道编码和调制器在某种程度上有相同的任务，即干扰保护。因此将两个局部的最优值整合成一个整体的最优值是有意义的。已经提出了一些方法（如编码调制、

栅格编码)，但它们仅对系统有了微弱的改善，因为信息论的极限最优值几乎已经达到。

2）传输系统仅对一定的信道特性是最优的。使用可适配（自调节）方法可以实现一个统一的系统。其中一个有意思的问题是如何识别信道（即测量信道特性）。

3）研究工作已经从近乎完美的点到点传输转到电信网络上。关键词包括网络组织（检测、控制、中断后的恢复等）和多接入。

我们也能看到理论由确定过程转为随机过程。与研究相关的电信工程也从电子工程变成统计学和计算机科学。目前给从事数学的人讲一些必要有关的电信技术的知识比给搞电子工程的人讲一些数学方面必要的知识还容易。

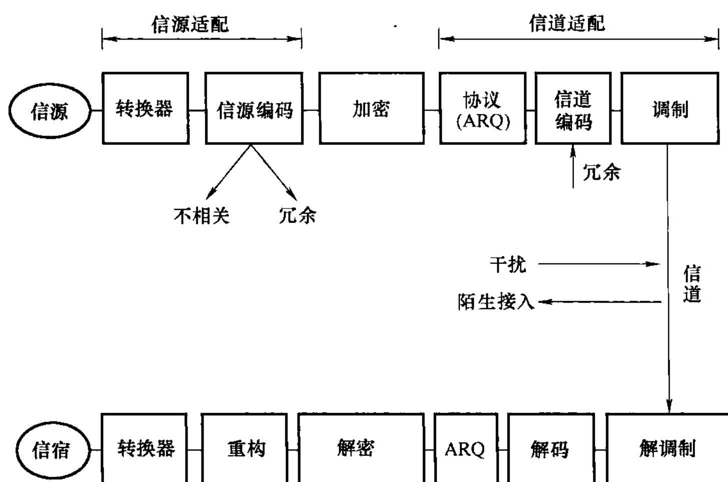


图 1-6 现代传输系统的完整组成结构图

总结：电信技术包括两个基本任务：传输之前和之后的信号扩展（编码和调制）；对传输本身进行控制（网络组织）。

1.1.3 数字信息传输的优势

“用数字的方式做！”这个概念已经渗透到整个信息技术中，这为下面的解释提供了一个很好的原因。

在模拟唱片上信息是以凹凸波纹的方式保存的，唱片表面的损坏会产生另一种波纹。对唱片播放机来说，它不可能知道一种波纹是本来就是这样的还是被损坏的结果（如灰尘、挤压、振荡)。

在模拟信号传输中每个信号的扭曲都会导致对信息的篡改。

调制（例如通过频率调制来增大带宽）是一种平缓的帮助方法。解决这个问题的主要方法还是数字技术！CD 同样是对光盘表面进行改变而存储信息的，但接收器（CD 播放器）只检测是否有孔存在。孔的形状是无所谓的（当然有个极限）。如果不小心将 CD 划了一下，这些小孔会因此出现磨损，但孔的存在与否是跟原来一样能检测

到的。但模拟信号在被数字系统处理之前必须要进行数字化。这由 A/D 转换器来实现，主要分成三步来完成：采样、量化和编码。

图 1-7 所示为一个模拟信号的采样。采样意味着以一个确定的、通常为常数的时间间隔对一个信号进行取值（采样值，英文为 sample）。采样时会得到信号的准确值。采样后得到的是时间上离散，但数值上是连续的信号。但这还不是数字信号！接着还要对采样信号继续进行处理。

现在可能有人会注意到了，采样后信息出现了丢失。又是香农发现了并不一定都会出现信息失真。他的著名的公式是

$$\frac{1}{T}=f_{\mathrm{A}}>2B \tag{1-2}$$

采样频率必须大于模拟信号带宽的两倍！

如果这个条件（即采样定理）被满足的话，采样值所包含的信息与原来的模拟信号所包含的信息完全相同。

香农公式将采样频率与带宽联系到了一起。带宽实际上是衡量信息量的一个维度。如果一个模拟信号中包含许多信息，那它相应地需要许多采样值，即需要较高的采样频率。如果一个信号包含较少的信息，它的带宽就小，信号变化得就较慢，采样间隔可以做得很大。

如果模拟信号是一个例如电子测量到的室内温度，采样意味着每 10s 测量出一个信号值。由于携带热量的关系室内温度不可能每毫秒变化 20℃。从感觉上看 10s 的采样间隔就足够了。

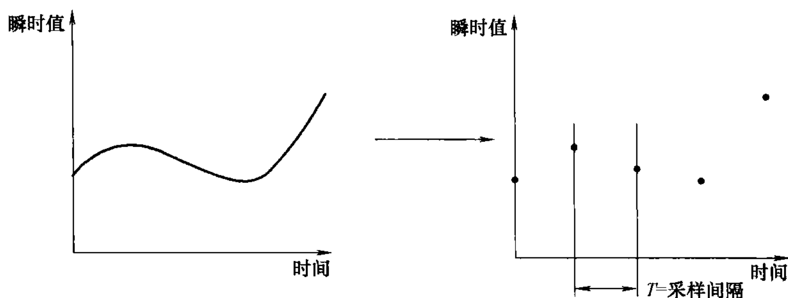


图 1-7 模拟信号的采样

比较图 1-7 的右半部分，可以看到不用原始的模拟信号就能用铅笔重构出信号曲线或对它进行插值。唯一的前提条件就是模拟信号在两个采样值之间没有跳变。这种快速的信号变化由于带宽的限制是根本不可能出现的。

音乐信号的带宽是 20kHz（可能更宽，但人耳是听不见的），CD 的采样频率是 44.1kHz。

采样之后是量化，如图 1-8 所示。这实质是对一定的离散值进行四舍五入，此后

信号不管是时间上还是数值上都是离散的，因而变成了数字信号。数字信号可以有例如 10 个可能的瞬时值，也可以有 17 个或仅有 2 个。后者被称为二进制数字信号，它在技术上是很重要的。在传输技术中还使用三进制（三个值）、四进制（四个值）信号等。最少需要两个可能的信号值：如果仅有一个的话，信号就变成了一个常数，即是确定的，因此它不包含任何信息量。

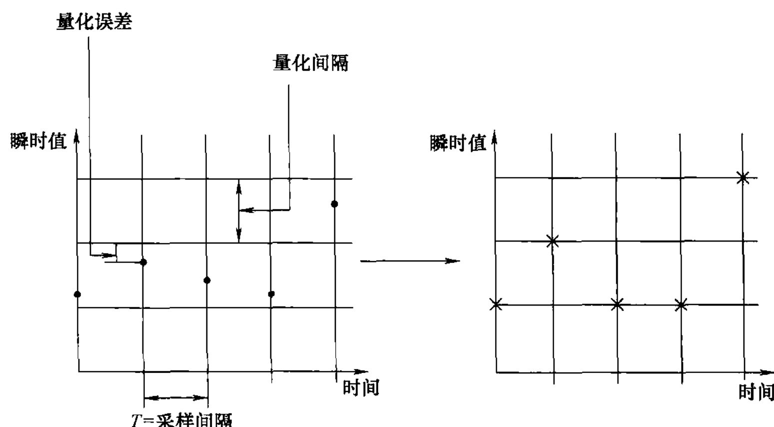


图 1-8　图 1-7 中时间离散信号的幅度量化

经过四舍五入的信号会发生扭曲（即量化误差），由此会产生信息丢失。量化误差最大是幅度量化间隔的一半，它表现为叠加在原始信号上的噪声，即量化噪声。这种信息丢失是找不回来的，是为了利用数字技术的优点而付出的代价。图 1-8 右侧给出了一个数字信号。它可以用一个数字串（序列）来表示，在我们这个例子中就是 [1，2，1，1，3]。因此信号就变成了计算机可以处理的形式。

不足之处当然是已经出现了量化噪声，目前可以通过减少量化间隔的方法弥补这种不足。如果划分幅度值的精度提高 10 倍，结果为 [1.2，1.8，1.3，1.2，2.7]；精度再提高 10 倍，结果为 [1.24，1.83，1.34，1.24，2.68]。原理是很显然的：量化过程中丢失的信息越少，数字信号所包含的信息就越多，当然描述后者所需要的数字就越多，信号处理和传输的复杂性就越大。

> 数字信号有一个离散的参数和一个离散的值的范围。在有限的定义范围片段内数字信号只包含有限多的信息。

这个定义有意识地让它适合所有的情况，但遗憾的是它不够简洁。两个三进制信号的例子应该可以将这些解释清楚：

1）一个时间和数值都离散的电压信号可以用一个序列（数字串）来表示或是像图 1-9 上那样用图形的方式来描述。从信息量上来说数字序列与图 1-9 中描述的是等价的，因为这两个图都是采用一个固定的规则，是可以相互代替的。

2）一个经过频率调制的数字信号被称为 FSK（Frequency Shift Keying，频移键控）信号。它是一个幅度为常数的连续曲线。有趣的是这个信号的频率，它在一定的时间内发生变化并只能取一定的值。尽管 FSK 信号的电压曲线是连续的但它却是一个数字信号，因为它关心的值（信息）不是幅度而是频率，如图 1-9 下所示。如果将频率曲线表示成时间的函数，结果就是图 1-9 中所示的阶跃信号。

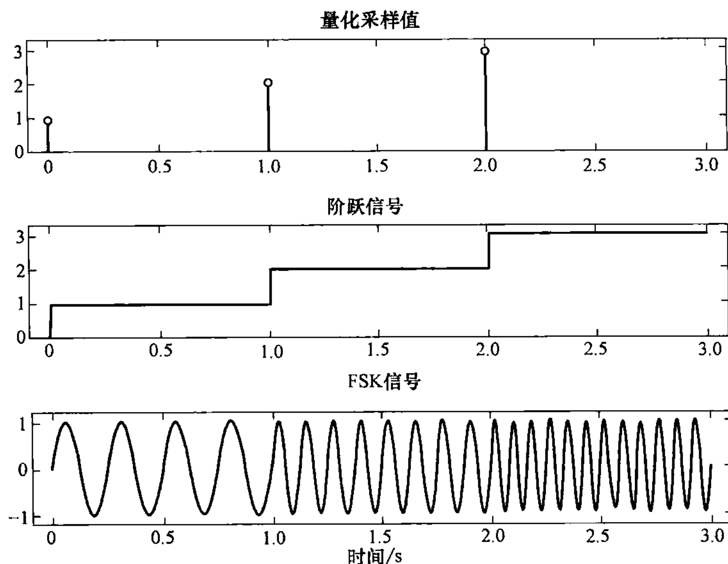


图 1-9 三进制数字信号

每条消息都是离散的！这种说法终于为有限的概念提供了理由。也可以这样说：足够细的量化（及其相应的复杂性）使每个信号都能得到足够精确的表达，即量化误差是可以忽略的。例如一个音频信号不必很精细地描述，以至于在耳的听觉门限值之下的空气分子热噪声都能够构造出来。实际中原始的音频信号也包含一定的噪声，它们不一定也要被数字化，因它们并没有描述任何有用的信息。这种情况下，量化过程中的信息丢失可以认为是减少不相关，即可以理解为是希望的数据压缩。因此可以夸张点说：

理论上，所有的模拟信号都可以被足够精确地数字化。

遗憾的是这只是理论上的！问题不是在量化上而是在采样定理上。实际中还没有足够快的 A/D 转换器和足够快的计算机，例如可以将雷达信号数字化。但是几百兆赫兹的采样频率还是可以实现的，例如图像处理已经可以实现数字化。

绝大多数情况下由数字单元处理二进制信号。假设信号电压是 0V 和 5V。传输误差（例如衰减）可能会使电压变成 4V，因此同样也把它当成 5V 来处理。甚至可以简单地将它再生成一个 5V。但当电压幅度是 2.5V 时就会出现问题，因此这是个禁止的范围。如果再生数字信号的时间过早，这完全是可能发生的。但对模拟信号来说这就不是问题，因为每次信号处理都会给信号增加噪声，在这之后都要把它跟有用信号一

样来处理。这些使用很小的一个实例就能证明：一首音乐被录制在一盒音乐磁带上。利用这盒磁带再制作一盒磁带，然后用制作的盒带再复制一盘，最后只剩下噪声，所有的信息都没了。用计算机磁盘做相同的事情，尽管使用相同的物理存储原理，但质量绝不会变坏。信息是免疫的，因为它可以再生，如图 1-10 所示。这种再生就是量化的结果。

数字信号可以被无限远距离地传输而不篡改信息的内容。

但再生器之间的距离必须足够近，否则当然是不可能。

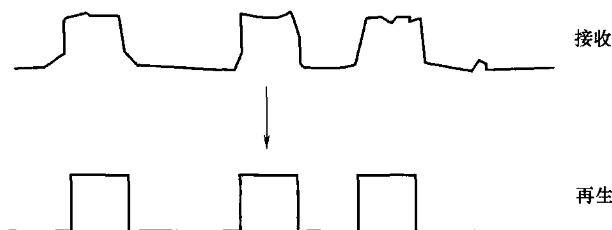


图 1-10 变型后的数字信号的再生

如果量化误差不应该太大，那么必须进行更细的量化。这意味着例如数字信号中包含着 1000 个值，如果要将它在 10V 的范围内划分，允许的值之间相差只有 10mV，一个幅度为 5mV 的干扰就能改变信号。但在上面提到的二进制数字信号中超过 2V 的干扰幅度才能改变信号。

同样的情况可以用另一种方式解释。发送器发送一个符号（例如一个电压值）到接收器中，接收器必须确定（以及估计或猜测）发送器发出的是哪个符号。模拟信号有无限多个可能的符号（例如电压值），因此正确的选择几乎是不可能的。精细量化后的信号也有多个可能情况，选择是很困难的。粗略量化后的信号只有很少的几种可能性（极限情况是两种），选择是简单的并且是可靠的。例如可以将 10 个数字想象成可能的符号（精细量化），出现传输误差时（没写清楚，或纸被折叠或湿透了），1 和 7 或 3 和 9 是很难区分的。如果只使用二进制码，只传输符号 0 和 1（粗略量化），那么区分起来就会变得很简单。

这样就陷入了进退两难的境地：一方面量化要尽可能精细，即有许多量化等级，这样量化误差才会尽可能小；另一方面，量化要尽可能粗，只有量化等级较少，传输的可靠性才会提高。这种进退两难的局面可以通过编码来解决。例如在图 1-8 中就使用了十进制编码。此时接收器必须要能区分 10 种不同的符号。如果使用二元码（二进制码）来代替十进制码，那么只有两个不同的值，传输的可靠性就会提高。但精细量化时会需要很多位，这在图 1-8 中已经很明显地可以看出。表 1-3 将十进制码与二进制码进行了对比。

符号多样性与传输可靠性的问题在文字上也会遇到：拉丁文比有上千个字符的日文要安全得多。也是基于同样的原因，传真是在日本而不是在欧洲发展起来：在欧洲人们觉得电传打字机很好，但同样的系统对日本字是不能实现的。传真时一张纸被分成了许多点，它们的值只是“黑”或“白”。这是一个二进制码，传输起来很可靠。

当然传输的许多黑-白值都需要判断，因此需要一定的处理速度。对机器来说快速地处理是很容易的简单过程，但对人来说正好相反。

表 1-3 十进制码与二进制码的比较

十进制码	二进制码
0	0
1	1
2	10
3	11
4	100
…	…
…	…
100	1100100
很多符号,较少位数	较少符号,较多位数

> 数字传输就是每秒钟对有限多的符号进行交换，符号只能有有限多个值。因此，每秒钟传输的信息量同样是有限的。

> 二进制码适合可靠的传输和机器的处理。

总结：

> 模拟信号的数字化（A/D 转换）由下面三个步骤组成：
> 采样：采样频率由模拟信号的带宽来决定。
> 量化：解析度由应用（精度要求）来确定。
> 编码：数字值通常由二进制码来表示。

声音信号的处理与文字信号相同。这种方法称为脉冲编码调制（Pulse Code Modulation，PCM)，它具有很广泛的应用。新的数字电话网络同样应用的是 PCM，即采样频率为 8kHz（话音带宽为 3.4kHz)，解析率为 8Bit（会产生 $2^8=256$ 个量化等级)。结果是每个话音信道的数据速率为 64kBit/s。

二进制数字技术有很大的优越性，因为它需区分两个信号值（符号)。

1）能够完整地重新生成已变形的信号。

2）数字信号中的信息具有免疫性。模拟唱片中音乐藏在波纹的偏转中。声音较小的部分偏转较小，声音较大的地方偏转较大。如果波纹变形了（例如由于灰尘颗粒的影响)，它对声音较小的轨道的影响比对声音较大的轨道的影响严重。相反在 CD 中，音乐是以凹点（孔）相当于位图的方式存储的。声音小的地方使用另外一种孔的分布方式。不论所表示的声音的大小如何，孔本身的形状都是相同的。因此声音较小的声道在输出中也不会被损坏。这在技术上被称为高动态范围。

3）复杂的、多级的处理步骤可以应用到数字信号中，同时可以保证误差不会像在模拟技术中那样变得过大。理论上复杂的算法也可以用到模拟技术中，但在实际中这

样的想法经常会失败（例如这会使人想到音乐盒带和计算机磁盘）。

4）信号处理单元可以不够精确，例如晶体管的电流增益可以在 100 ~ 500 之间波动。可以用不精确的单元构造整个系统。

5）使用高集成技术是可行的。

6）所有类型的信息都用同一种方法进行描述，即比特序列，不管是音乐、话音、图像、文字、控制命令还是账户信息。因此最先有了 ISDN（Integrated Service Digital Network，综合业务数字网），即所有的信息都经过同一个网络传输。

7）数字技术是统一的、各种环境下都能使用的技术。各种单元可以大规模地生产，高集成的巨大的开发费用可以分散开来。微电子技术从经济角度上来说变得是可行的。

8）数字技术使用计算机处理成为可能。因此处理可以使用软件控制的机器来进行。设备的修改意味着修改软件。此外仿真实现成为可能。

9）光波导（光纤）特别适合传输数字信号。几 GBit/s 的信道容量是可以实现的。高性能的、用于数字信号传输、交换和处理（即能完成所有电信任务）的系统已经存在。

> 数字技术提供了一种很好的功能，它具有很高的灵活性和运行可靠性，它可以以很低的价格大规模地运行。

数字技术会统治现代电信技术是不足为怪的。模拟技术的新发展只能用在根据采样定理不可能数字化的地方（高频）或需要较高功率的地方（发送端功率放大器）。

1.1.2 节中已经介绍过了，采用频分复用的方法进行信道复用是一种很经济的传输方式。公共交通工具的原理也能用在数字信号上：

例如当一个用 1kHz 采样的信号使用 4Bit 进行编码，脉冲宽度最多是 0.25ms。但它们可以更短，例如 0.025ms。多个这样的短脉冲数据流可以混合在一起传输。同步系统使接收器中数据的正确排序成为可能，如图 1-11 所示。这种方法被称为时分复用（Time Division Multiplex，TDM）。它用在长途电话网中，当数字速率为 2.5GBit/s 时大约 30000 个话音信道可以使用一个光波导来传输。

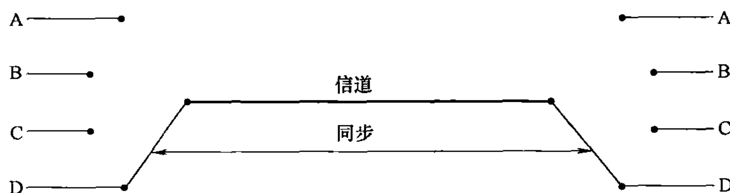


图 1-11　时分复用方法的原理

1.1.4 信息论基础

为了理解电信技术还需要一些信息论的基本知识。读者从下面的介绍中应该找到有关信息论的一点感觉。但这里没有给出有关数学方面的内容，因为定量部分这里还不需要，4.1.1 节将会给出准确的描述。

在给信息定义时人们会遇到这样的困难，即 Information 这个说法的口语含义对电信技术来说完全是不可用的。电信技术需要的是一个没有值的定义，能够对信息本身进行量化（即数量描述，不要与量化弄混！）。相反信息的内容量、实时性和值是主观评价，在技术的、纯量化的定义中没有什么值得寻找。在一幅画中，对不同的观察者而言，可能线条或是颜色是最重要的。但对一个传输系统而言，接收者的评价一点不起作用。接收者的评价对信源编码来说是起作用的：在某些评估是已知的情况下，可以通过减少不相关来实现数据压缩！另一种比较合适的对信息的描述方法是：

> 信息就是克服不确定性。
> 对接收者来说信息就是新颖性。
> 信息有令人吃惊的瞬间。

这里并没有说，接收者是否愿意接收信息并且能处理它。根据这种描述，很显然常数信号不包含任何信息，同样周期信号运行一个周期之后也没有任何新颖性。人们可以在接收一个周期后将信道阻塞，然后在转换器的输出端（见图 1-3）对接收到的信号进行重复，信宿看不出任何区别。这对任何确定信号都是可能的。

> 只有随机信号才含有信息！

严格说来没写入信息的常量信号根本不用传输，从 $t=-\infty \sim +\infty$ 它都是存在的。周期信号也是这样的。如果为了数学描述也允许使用这些“永恒的”信号，那么这个理论就会变得很简单并且很紧凑。同样有无限大值的信号在数学描述中也被使用（如 Dirac 冲击），尽管没有系统能处理它。

在所有的通信中都是交换特定的、具有事先约定好的含义的符号。符号可以是例如字符（光符号）声音或是音素（声音信号）。通常情况下一整组符号有一个含义，人们将这个组称为字。书写的字符组合“Haus”和发出声音的“Haus”在接收器中会丢失掉相同的相关性。因此说外语的人根本无法开始，因为事先没有约定。例如德语中说的词“Igel”对美国人来说是一种食肉鸟，而不是刺猬。同样做为信源和信宿的机器也需要一种通信约定：例如对光信号设备的控制器给交通信号灯发送一个位图，现在“1”意味着灯亮还是灯灭？两种情况都是可行的，通常情况下两者也具有相同的含义（顺便提一下，在光信号设备中正好相反：中断控制和信号灯，后者获得与逻辑“0”系统的符号，在这种情况下不能简单地让绿灯亮！）。

因为通信约定只是在信源和信宿之间一直存在的有限的概念上的术语，因此：

> 信息传输是符号的交换。符号来源于符号池（即字母表）并且具有实现约定好的含义。

信宿的概念是有限的，因为它不可能比能感觉到的外部世界的解析能力（区分能力）更大。例如：光信号的颜色由光的频率来决定。这个频率可以无限地变得更精细。如果提出这样的问题，人到底能区分出多少种红颜色，其实答案的数量很少。即人眼的频率解析能力是有限的。数一下实际中使用的描述不同特性红色的词汇，这个数量

就更少了。红色的主观含义比传感器能够感觉到的更少。

人有一个特殊能力，即不依靠传感器可以感知外部世界，从“自己内部”就可以产生信息（例如想法）。人也会使用抽象的概念。在这点上一个重要的任务就是产生语言，将语言作为这些抽象概念的载体。当概念的数量有限时，人的特殊角色并没发生改变。

符号可以借助于信号的帮助以任何的物理形式传输。如果接收到的信号能将符号正确地组装，那么通信就是无误差的。为此又出现了符号的多样性与传输的可靠性之间的矛盾，解决方法就是编码：将有限多的符号被编号，有目的地选择了二进制码。编码无外乎就是字母表的变化，例如字长为 1，在字母表中顺序为 32 的码转换成字长为 5，在字母表中的顺序为 2 的码。后者就是二进制码或二元码。

用来传输的物理信号一直被认为是模拟信号，它可以表示无限多的符号。因为这之中只有有限多的部分被真正地用到，因此可以实现一定的抗干扰功能。相反对理论数学来说这通常指的是有一定目的的数字信号。因此信号描述的复杂性与信号所包含的有效信息量的复杂性一样大。

一个数字信号仅仅是翻译成数字的模拟信号。

所有的信息最终都是数字的。 所有的物理信号最终都是模拟的。

因为现在信息是可数的并且可以用二进制码来描述，因此使用 bit 作为信息量的单位（它来源于英文 binary digit）。

关于写法的说明：小写的 bit 是信息量的单位，而大写的 Bit 表示二进制码的一个符号。一个 Bit 最多可以表示 1bit 的信息量，通常情况下少于这个值。

每个概念必须是符号或字（即符号的组合）。为了提高符号之间区别的程度，要最小化不同符号的数量，理想值是只有两个符号。因此才有了前面已经提到过的编码。由此出现了数据字，正是由于有了通信约定它才变得为大家所知。在书写语言中字母是作为符号来使用的，由这些字母组成了单字，这才有了概念。口语将声音（音素）组合成单字。

现在符号的组合可以表示一个概念，但不是所有的符号都相互无关了。例如德语的书面语言中“c”后经常会出现“h”。因此出现“h”的惊奇程度要比出现“c”的惊奇程度小。根据上面的定义“h”包含的信息比“c”包含的信息少。但这仅对德语有效，即对一种特殊形式的编码有效！如果不使用字母而使用二进制符号（Bit），那么它们之间的依赖性同样存在（冗余）。因此 1Bit 并不一定包含 1bit 的信息，绝大多数情况下要少于这个值！

如果所有的字母在德语中出现的概率相同，那么每个字母所包含的信息量为 $\log_2(26)=4.7$bit。由于字母间的相互依赖性，这个值变成每字母 1.6bit。这没有区分大小写，是正确的，因为大小写有固定的规则，并不影响惊奇程度，因此不包含任何信息。

> 符号出现的概率越大，这些符号所携带的信息就越少。

信源编码的目的是通过编码转换尽可能地实现极限值 1 Bit = 1 bit。一个 Bit 所包含的平均信息量被称为熵。在热动力学中这种方法表示统计无序。热动力学和信息论都涉及统计现象，因此两种理论有相似之处。减少不相关的信源编码也是减少熵。

语言中的相关性对字母组合也同样是适用的。字之间也有相关性。在语言中使用相似的字来表达不同的概念是没有危险的，例如 “Nacht”（夜晚）、“Macht”（权利）、“Yacht”（游艇）、“wacht”（监护）、“lacht”（笑）、“sacht”（轻轻地）。传输误差可以根据上下文进行纠正，但这需要在接收端有一个有识别能力的判决过程。简单的方法并不能识别出 “Macht” 是由于传输差错而导致的，应该将它纠正为 “Nacht”。但当出现 “Kacht” 时就比较简单，因为德语中并没有这个词（或至少作者不认识这个词，这正表明了通信约定的重要意义！）。遇到 “Kacht” 这个词时就可以立即判断出这是传输误码，聪明的人类可以根据上下文对它进行纠正，而比较笨拙的机器能报告出差错也足够了。人甚至可以用同一个词表达不同的概念，例如 Zug，这个词的含义（气流、火车、牵引力、瑞士一城市）当然只能联系上下文才能确定。

现在我们要弄明白，信道编码应放在何处：可选的符号越少、符号之间的区别就越大，接收器对各个符号的区别就能做得越好。因此传输数字信号（只有例如两个符号）尽管也会遇到差错，但笨拙的机器看到的是尽可能不相似的字（Bit 组合）。信道编码器的作用就是将相似的字变得不再相同。例如根据 “Macht” 产生出 “ABCMacht”，根据 “Nacht” 产生出 “XYZNacht”。这样 M 和 N 的差错不用改变上下文就能被发现，甚至可以被纠正。目前已经有一些很好的数学方法，能找到很合适的将字加长的方式。这些附加的字母或 Bit 当然并不携带任何新的信息，因为它们与当前这些较长的词的其余部分是紧密联系在一起的。它们只是对已有的信息的重复，即它们是冗余的。尽管如此还得传输它们，并且消耗时间和/或带宽（见式（1-1））。尽管有信号的干扰并为此而进行了符号转换，所得到的回报是减小了信息对干扰的敏感性。接收器中的解码器（见图 1-6）必要时可以对接收到的字进行纠正，去掉冗余（“ABC” 以及 “XYZ”），并将信息继续传递。字母的拼写（“Anna”、“Berta”、“Carlo” 等）也是信道编码。

到目前为止我们只谈到了信息数量，现在我们也要看看信息含量。以此为目的的一个模型将信息内容分成了 4 组，如图 1-12 所示。只有设计巧妙的、感兴趣的部分是必需真正要传输的。信源编码的目的就是将其余的部分去掉。

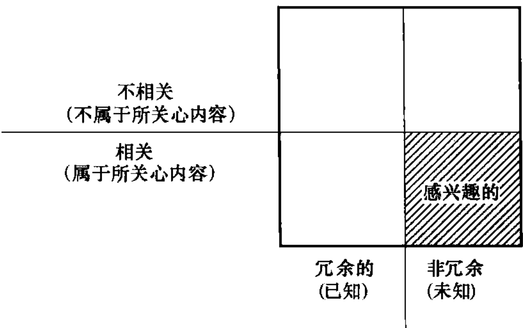


图 1-12　信息的分类

区分相关-不相关与应用有关，因此是主观的。没有人会将报纸一

字不漏地从头读到尾，多数情况下读者都是找与他相关的部分来读。不同的两个人可能会选择完全相同的部分。因此很显然他们不论是相关还是不相关的划分都是相同的，即分成相关部分和不相关部分。

理想的信源编码器会将整个不相关部分都消除掉（例如广播新闻报道中的环境噪声）。这是不可逆的信息丢失（熵减少），但这样做并不心痛，因为否则的话信息就变成相关的了。A/D 转换器中的量化也可归结为减少不相关：如果量化间隔小于信宿的解析能力，量化噪声根本就不存在，因此就只是减少不相关。

信源编码器也能消除冗余，但这是个可逆的过程。

在人类的语言中，相关和不相关之间的界限大不相同：仅仅是内容上，频率范围就从 700 ~2000Hz。再加上说话者还分男、女，因此频率从 300 ~ 2000Hz 是必需的。为了能辨别出来说话者，必需要传输 300 ~ 3400Hz 的频率范围，这与正常的电话信道是一致的。如果要想将一首歌很逼真地恢复过来，大于 5kHz 的频率是必要的。从式（1-1）中可以很明显地看出，高带宽意味着高信息量。

话音传输所需要的带宽是根据对音节的理解经过长时间的尝试得到的。测试人员大声阅读毫无意义的音节（所谓的 Logatome），然后测试能够识别出的比例。采用这种方式能够阻止根据声音大小之间的关系而下意识地进行重构。当接收比例很差时，使用外语文字比使用母语文字难懂很多。这可以简单地用收听 Pop songs 来证明。音节理解能力（根据 Logatome）为 80% 时，一句话（上下文是有意义的）能被正确理解的概率为 96%。这种比例上的区别在于上下文是有冗余的。80% 的音节理解率可以通过 300 ~2100Hz 的信道来达到。如果使用频率范围为 300 ~ 3400Hz（跟电话一样）进行传输，那么音节理解率可以提高到 90%。

冗余和不相关的区别通常是指所要求的传输质量的问题。这个问题不能简单地用“尽可能好”来回答。

信息的另外一种划分办法在语法、语义以及实用性方面。语法描述信息的组成（例如在语言中就是语法和拼写规则），语言描述它的含义，实用性描述对接收者来说信息的价值。信息论仅涉及语法方面。

现在我们再回到信息量中来！通过几个数值例子应该对数值大小的顺序找到感觉。

1）话音：电话质量的话音产生每秒钟 64000Bit 的数据量，相当于 64kBit/s 的数据流或传输速率，这在前面就已经提到过了，此时并没有使用有效的信源编码。话音的实际信息量（没经过润色，相当于一个最优的对文本的信源编码）视说话的速度不同仅为 5 ~25bit/s。

2）音乐：CD 能够达到非常逼真的声音质量。因此对最大声音频率为 20kHz 的音乐信号所使用的采样频率是 44.1kHz。所有的值被分成 $2^{16}=65536$ 个中间值，这些值用字宽为 16 个 Bit 的二进制码来表示。总共产生的数据流为 $44000\times16\approx700$kBit/s。在 CD 中用两个音道（立体声）来存储，最长的存储时间为 75min。因此总的信息量为 6GBit。除此之外还要加上信道编码和复杂的同步系统、音乐曲目的编号、时间等。

3）静止图像：是衡量研究能力的一个尺度。整幅图像被分成许多点。眼睛的视角

约为 100°，视角的解析率约为 0.1°。在每个维度上大约能区分 1000 个点，两维的图像包含 $N = 10^6$ 个点。

眼睛能区分不同的亮度（灰度级别）。使用 256 个级别就能给人很逼真的亮度感觉，这些级别需要用 8 个 Bit 来表示。人眼还能区分颜色，可见光光谱的宽度约为 2500Å，人眼能够识别的频率解析率为 10Å，因此能够区分出 250 种颜色，为此还需要 8Bit。

4）活动图像（视频）：每秒钟传输 30 幅图像是必需的，这样眼睛才不会感觉到图像的闪动。这样数据传输速率大约是 500MBit/s。

在我们的眼中看起来很完美的图像传输需要的信息速率为 500MBit/s。对图形的评价是在大脑中进行的，那里有大于 10^{12} 个神经元（存储位置）。大脑需要的纯计算需要 2000s，即经过半小时之后就会填满了信息！

这个例子表明，我们的传感器和大脑能完成不可能的减少数据的任务（除了能很巧妙地组织存储和不忘记之外）。人类有意识的信息获取的速度仅为 20Bit/s，正好相当于上面所提到的语言的最大信息量。因此原则上人类对信息的需求由铜线来提供是可行的，尽管目前比铜线容量大上百万倍的光波导已经投入使用。

很显然人类已经能够从所供给的信息中通过消化得到整体印象，而不被细节所误导，但这个选择是根据主观的评价。这正是虚拟现实的主要问题：我们要选择相关的数据，使人在计算机世界中感觉就跟在现实中一样。那么这样信息处理的计算复杂度是否还停留在可实现的范围内？另外一个问题是参与者（即信号转换器），在一个很重的数据头盔下，谁还会感觉很好呢？

将一幅正常电视图像数字化，需要 140MBit/s（黑/白）以及 216MBit/s（彩色）的速率。跟上面提到的 500MBit/s 做比较，可以看出已经有了一定的质量损失，例如每秒只传 25 幅图像并且每幅图像只传较少的点。这涉及的是减少不相关。使用巧妙的减少冗余的办法会使数据传输速率继续下降到 34MBit/s，并且观察者并不能感到任何区别。在几乎感觉不到质量下降的情况下甚至 5 ~ 10MBit/s 的数据传输速率也是可能的。表 1-4 给出了一个对数据传输速率的总结。

表 1-4　不同信源信号的数据传输速率（指导值）

信号内容	未经压缩的数据传输速率	经过压缩的数据传输速率
话音（单声道）	64kBit/s	2.4 ~ 32kBit/s
立体声音乐，CD 质量	1.4MBit/s	128kBit/s
视频（黑/白）	140MBit/s	2 ~ 5MBit/s
视频（彩色）	216MBit/s	5 ~ 10MBit/s

1.1.5 电信技术中的系统工程

整体大于个体的和！一个系统具有这样的特性，即只根据单个部件不能推出整体的性能（例如：Lego 的拼装玩具）。系统是具有内部组织和一些外部可见特性的执行

联合体。

系统理论方法是在工程领域中应用很广的一种工作方法。人们使用一种综合的思考方法（根据在所处环境中的作用来定义系统），跟分析描述方法（根据组成单元进行系统定义）刚好相反。这样做会产生两种后果：

1）只对系统的功能（作用，输入的大小对输出大小的影响）感兴趣，对它的内部结构不感兴趣。整个系统被描述成一个黑盒，如图1-13所示。

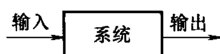


图1-13　作为带有（多个）输入-输出黑盒的系统

2）整个系统可以被划分成几个子系统。系统是分层次构成的，每一层都使用相同的工作方法。子系统又可以看成是一个黑盒子，这些子系统的功能连在一起就构成了整个系统，如图1-14所示。

子系统的构成以及分级组成的目标是将复杂的功能一步一步地完成。要避免从上到下来对系统进行抽象，否则细节程度会上升。上一级描述与系统应用相关的内容，相反下一级描述与实现相关的内容。

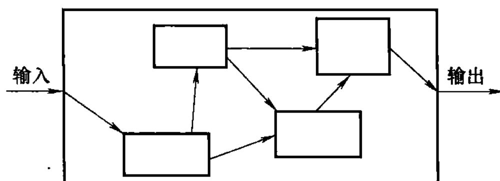


图1-14　带有子系统的系统

整个系统的特性由各个子系统共同来实现。以较大的复杂性来改善每个子系统，以期望整个系统的性能能得到自动的改善是毫无意义的。

系统的设计以用户的需求为出发点，遵循自顶向下的方法。用户的需求被翻译成技术数据，这些数据不一定是紧凑的。然后划分成子系统，每个子系统采用相同的方法。这些工作很具有挑战性，大大超出这些简短的描述所给人的印象。一种不恰当的系统设计通常很晚才能被发现（有时根本未被发现），通常产生的结果是功能描述的最后一点点实现要付出不相称的巨大并昂贵的复杂性。

系统应该如何划分是自由决定的（也是由经验决定的）。适当的划分能使分离子系统的开发和测试成为可能。下面的规则对系统的划分是有帮助的：

1）只有对当前的问题有意义的那些信号才应该作为输入和输出出现。

2）输入大小的选择应使它对系统的过去不会产生影响。

3）困难的任务应完全包含在一个子系统中或完全消除。只有在层次划分得很多时，一个棘手的任务才应该分到两个子系统中。通过这种方法不必寻求成本较高的局部最优。

4）一个子系统应该是地理位置紧凑的本地可实现的。

使用系统论方法的好处有以下两个方面：

1）在复杂系统的描述中，只有与目前要解决的问题相关的方面才是可见的。人们看到的是森林，而不是树木。

2）实施的考虑与实现分离开来。同样的方法可用在技术、生态环境以及社会系统中。

一般性的讨论之后，现在应该将这个概念用在电信传输系统中。图1-15所示为从应用者的角度看到的最简单的系统描述。

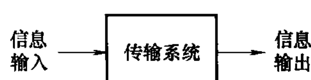


图1-15 站在应用者角度上的传输系统

对技术人员来说还必须考虑各种输入。图1-16所示为一个短波话音无线传输时的例子。

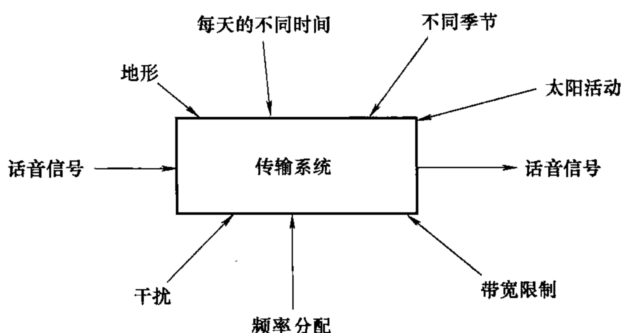


图1-16 使用短波传输话音时对图1-15的细化

下一步是将图1-16所示的系统划分成几个子系统。参见图1-17。它要说明的是多个输入能分到一个子系统中。接下来可以根据图1-16对发送器和接收器进一步细分。

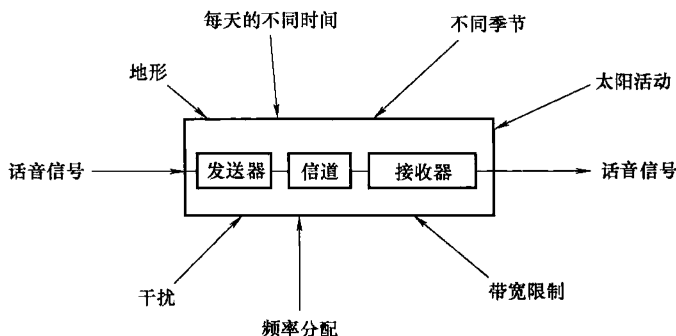


图1-17 图1-16的细化

但接下来的划分并不是很相近的，需要自己来判断。例如：发送、接收天线属于发送器、接收器还是信道？两种可能都有意义，两个天线都可以属于或都不属于信道。哪种方法有优势与问题提出的角度有关。通常情况下天线，包括它们的馈线都算在信道上，因为这样渗透到系统中的是电压和电流而不是电磁波。

信道的定义也是自由决定的，这点在1.1.7节中还要解释。

经过概要介绍后是这样的。目前系统工程是很重要的（多面手是很受欢迎的!），因为许多复杂的系统，例如数据网不是一接上就能工作的，尽管所有的设备都是无错

的。这种情况下寻找错误需要大量的系统知识、有组织的过程以及坚持不懈的努力。相反对设备的了解意义并不是很大，当然这也是无妨的。

1.1.6 基准和衰减

信息传输同时也被称为能量传输。这在式（1-1）中已经表现出来了，那里先出现的是信号功率和传输时间。信号在传输介质中的衰减也必须仔细进行考虑。

现在要将传输信道作为一个二端子来考虑，它是由如传输线路、放大器等那样的线性四极子级联来实现的，如图 1-18 所示。

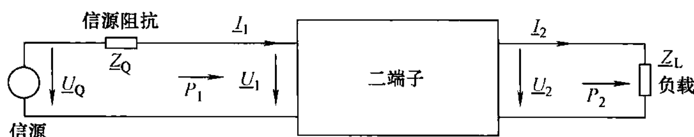


图 1-18 二端子作为发送器和负载之间的信道

利用线性二端子人们可以用复杂的标记方法。放大定义为用复数表示的输出端谐波大小与用复数表示的输入端谐波大小的比值，这正好相当于四极子的频率响应 $H(j\omega)$。作为公式的符号除了 H 之外还有 G（增益）和 V（放大）。为了方便起见尽管这些符号是复数，但通常不加下划线。衰减 A（Attenuation）是 H 的倒数，因此同样是个复数值。

电压放大：
$$\underline{V}_u = \frac{\underline{U}_2}{\underline{U}_1} \tag{1-3}$$

电压衰减：
$$\underline{A}_u = \frac{1}{\underline{V}_u} \tag{1-4}$$

电流放大和衰减：
$$\underline{V}_i = \frac{\underline{I}_2}{\underline{I}_1};\ \underline{A}_i = \frac{\underline{I}_1}{\underline{I}_2} \tag{1-5}$$

功率放大和衰减表示的是有用功，因此是实数。

功率放大和衰减：
$$V_p = \frac{P_2}{P_1};A_p = \frac{P_1}{P_2} \tag{1-6}$$

所有信号在传输的过程中由于被吸收而产生衰减。这其中在信道中存储的功率的一部分转换成了热量。但不能将所有的损失都用瓦特来进行绝对的衡量，而要对它进行相对的（例如用百分比）量化。因此所有的能量吸收过程都可用指数函数来描述，信号幅度随传输的距离的增大非线性递减，如图 1-19 所示。

在实际使用中若幅度随着传输距离的增加而线性下降，那么会比较方便。这可以通过对幅度轴用 e 函数即对数函数的倒数进行变换来实现。这导致了单位奈培（Np）的出现。

当幅度值的范围超过几十的时候，也习惯使用对数刻度。通过取对数，数值的范围会被压缩而在整个范围内的相对精度会保持不变。通常情况下使用的单位是分贝

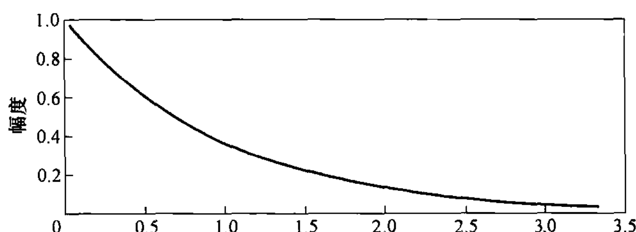


图 1-19　有损信道信号幅度与路径长度之间的函数关系

（dB），它是底为 10 的对数。最原始的定义是从功率出发的：

$$V_p[\mathrm{B}] = \lg \frac{P_2}{P_1}$$

这个 Bel 从来没被用过，但是它的十分之一，却变成了分贝：

功率关系：
$$V_p[\mathrm{dB}] = 10\lg \frac{P_2}{P_1} \tag{1-7}$$

相反奈培变成电压比的自然对数来使用：

$$V_u[\mathrm{Np}] = \ln \frac{U_2}{U_1} \tag{1-8}$$

因为不同底的对数之间有一个常数的换算关系，dB 和 Np 之间存在一个简单的关系：

$$1\mathrm{Np} = 8.686\mathrm{dB}$$

$$1\mathrm{dB} = 0.115\mathrm{Np} \tag{1-9}$$

如果信号的幅度是以 dB 或 Np 为单位给出的，那么图 1-19 的曲线就是线性的了。图 1-20 所示为一个这样的功率图（这种方法的前提条件是电路没有反射）。

值得注意的是：

1）dB 是无量纲的实数值之间的比值，归一化会引起值的计算错误。

2）dB 最初是为功率之间的关系而定义的。它也经常被用在电流或电压之间的关系上。此时存在如下的关系：

$$\frac{P_2}{P_1} = \frac{U_2^2 R_2}{U_1^2 / R_1} = \left(\frac{U_2}{U_1}\right)^2 \cdot \frac{R_1}{R_2}$$

$$V_p(\mathrm{dB}) = 20\lg \frac{U_2}{U_1} + 10\lg \frac{R_1}{R_2}$$

当 $R_1 = R_2$ 时，上式的第二项就不存在了。只有在这种情况下，dB 用来表示功率比和表示电压比是相同的。

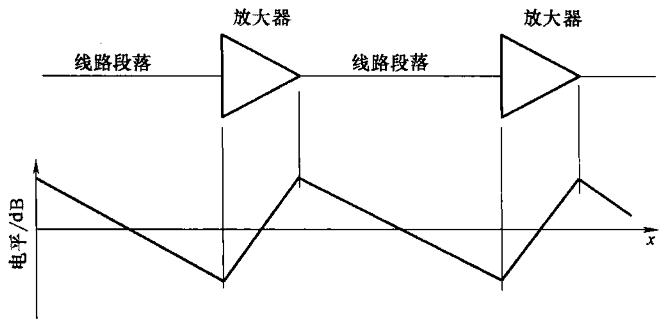


图 1-20 上：三段、两个放大器的线路连接
下：对应的功率图

电压比：

$$V_u(\mathrm{dB}) = 20\lg \frac{U_2}{U_1} \tag{1-10}$$

功率比：

$$V_i(\mathrm{dB}) = 20\lg \frac{I_2}{I_1} \tag{1-11}$$

当电阻值不相同时（传输路径中的改变），必需说明所涉及的是电压比还是功率之比。

尽管人们还必需习惯使用其他用法，但在实际应用中，使用分贝还是有很大帮助的。表 1-5 给出了几个分贝值，其他值可以简单地通过心算就能得到。

表 1-5 常见的分贝值

	功率比例	电压比电流比
20dB	100	10
10dB	10	$\sqrt{10}\approx3.16$
6dB	4	2
3dB	2	$\sqrt{2}\approx1.41$
0dB	1	1
-3dB	0.5 = 1/2	0.707
-6dB	0.25 = 1/4	0.5
-10dB	0.1 = 1/10	0.316
-20dB	0.01 = 1/100	0.1

使用无量纲的 dB 同样可以表示无量纲的放大系数。但我们感兴趣的还是实际的功率值、电压值或是电流值（绝对大小），所以必需约定一个基准值（式（1-7）、式（1-10）或式（1-11）的分母的大小）。人们将它们称之为基准（level L）：

电压基准：

$$L_u = 20\lg \frac{U}{U_0}\mathrm{dB} \tag{1-12}$$

电流基准：

$$L_i = 20\lg \frac{I}{I_0}\text{dB} \tag{1-13}$$

功率基准：

$$L_p = 10\lg \frac{P}{P_0}\text{dB} \tag{1-14}$$

绝对基准是国际上定义的基准值。表 1-6 所示为常用的基准值。通常人们直接用 U 或 I 或 P 来代替公式中的符号 L。因此 $U = -60$dBV 意味着 1mV，$P = 13$dBm 意味着 20mW。相反放大和衰减是无量纲的比值，单位总是 dB 而不是 dBm 等。

表 1-6 绝对基准的基准值

单位描述	基准值	归一化值
dBm	$P_0 = 1$mW	0dBm 相当于 1mW
dBW	$P_0 = 1$W	0dBW 相当于 1W
dBV	$U_0 = 1$V	0dBV 相当于 1V
dBμV	$U_0 = 1\mu$V	0dBμV 相当于 1μV

通常情况下伏特计使用 dBm 或 dBV 这样的对数值数量级就足够了。很明显 dBV 表示的是电压，而 dBm 表示外部电阻上所测量到的电压产生的功率。因此在第二种情况下必需声明基准电阻。通常情况下这个值为 600Ω（电话）或 50Ω（HF 技术）。

例如：

1）刻度：dBm（600Ω）：0dBm 相当于 1mW 的功率，600Ω 电阻上的电压值为 775mV。

2）刻度：dBm（50Ω）：0dBm 相当于 1mW 的功率，50Ω 电阻上的电压值为 224mV。

1.1.7 传输信道

传输信道是发送器和接收器之间的链接，如图 1-2 所示。发送器和接收器两者都能划分为多个子系统，参见图 1-3 和图 1-6。发送器的任务是将信号适配成适合在信道中传输，例如在无线传输中使用调制器对信号的频率进行搬移。另外发送器还得使用放大器准备足够的功率，以补偿传输中的衰减。接收器需要进行解调，即将信号的频率变回到原始的频率值上。

但也可将这些信号变换看成是信道变换。图 1-21 所示为上述无线话音传输例子的系统组成方块图。话音信号是基带信号，即它包含从低频一直到一定的极限值的频率成分。而无线线路只允许大于 100kHz 的频率值通过。因此在图 1-21 中可以区分成 3 种信道，每个发送器子系统要对下面的信道进行信道适配：

1）信道 1：具有较大衰减的带通信道；

2）信道 2：具有较小衰减的带通信道；

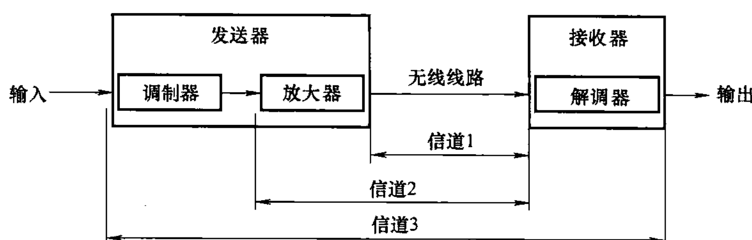


图 1-21 信道转换

3）信道 3：具有较小衰减的基带信道。

通常情况下这些信道是混在一起的，内部信道完全被外部信道所包含。

传输信道与传输介质要区分开来！

传输介质（有时也称为物理信道）包括双绞线（twisted pair）、同轴电缆、光波导（LWL）、无线线路、红外线和激光。

第 5 章将详细介绍这些介质。传输介质总是信道（当然通常情况下是最内层的），但反过来不一定！信道是个比较通用的概念。传输介质只传输物理（最终是模拟的）信号。没有数字介质，但存在数字信道。（这可以改变：在原子领域必需考虑到量子物理。当应用涉及到这些领域时，也可能会有数字介质！）本节的后面部分不讨论介质，而是讨论信道。

模拟传输信道由下面的指标来刻画：

1）频率范围：一定的允许通过的范围（频率下边界和上边界）。

2）允许通过范围内的衰减。线性信道情况下，大多数情况由是复数的与频率相关的传输函数或频率响应来描述，通常使用 Bode 图。

3）通过加法和乘法效应所引起的信号变形：

-固有噪声，如电阻噪声和半导体噪声（加法，不确定性的）；

-接收到的干扰信号（加法，不确定性的）；

-宇宙和大气的噪声、量化噪声、串扰、网络散杂干扰（50Hz 电磁干扰）、开关操作引出的散杂干扰（高压设施）；

-线性失真（乘法、确定性的、可以纠正的）；

-非线性失真（乘法、确定性的、可以纠正的）。

数字传输信道由下面的指标来刻画：用 Bit/s 表示的传输速率，误码率（Bit Error Ratio，BER），差错的类型（平均分布还是时间累积等）。

现在我们得仔细看一下信道的频率状况。信息信号（如音乐、数据）是人们实际想要传输的信号。信号的带宽与信息含量有关，参见表 1-1。频率下限或者是 0Hz（数据）或者是几 Hz（音频信号）。重要的是在大多数信息信号中频率的下限是固定的，即与信息量无关。因此通常称这种信息信号为低通信号（TP 信号）或通称为基带信号（BB 信号）。后一种表达更常见，因为音频信号严格说来并不是 TP 信号，0Hz 并不会

出现。通常在要求不是很精确的情况下两者会交替使用，根据上下文之间的关系他们新要表达的真正的含义会很清楚。

在实际实现传输系统时，不含 DC 分量的信号对于简化系统的实现有以下三大好处，这在模拟电话系统中已得到了充分的应用：

1）在传输路径中可以很容易地使用用于绝缘的变压器，这样可以避免平衡补偿电流（不同的接地电流）所带来的干扰。

2）在放大器中信号和工作点电压可以很容易分离。

3）（埋在地下的）中间放大器的馈电可以使用传输信号用的双绞线来同时完成。

对于传输来说，信息信号的频率范围（频谱）与信道允许通过的频率（允许窗口）相匹配是很重要的。如果不相匹配（例如用来传输 TP 信号的带通信道），那么必需使用调制器及解调器来进行信道适配。

需要对下面各点进行区分：

（1）BB 信道（基带）

1）DC 耦合（低通）；

2）AC 耦合。

（2）BP 信道（带通）

TP、BB 以及 BP 信号之间的区别是相同的。由于信道的物理特性中不存在高通（HP）或具有带阻（BS）特性的信道，这些信道的允许带宽是无限大的。

信道的带宽（允许通过的范围）是可变的，它与信道容量（信号的情况下是与信息量）有关。

如图 1-22 所示，BB 信道或 BB 信号时只有频率的上限与带宽有关。BP 信道或 BP 信号时频率的上界和下界都是可变的。

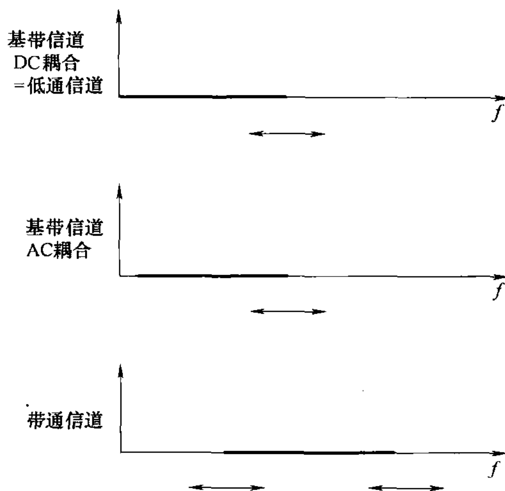


图 1-22 TP、BB 及 BP 信道和 TP、BB 及 BP 信号允许通过带宽的比较

最简单的情况是一个 TP 信号使用 TP 信道来传输，如图 1-23 所示。这种情况会采用 1.2 节所述的模拟信息信号和第 2 章所述的数字信息信号来处理。TP 信道指电缆和电线。如果需使用用于绝缘的变压器，那它则涉及用于 AC 耦合的基带信道。

无线传输时涉及的始终是 BP 信道。因为电磁波只有在大于 100kHz 的频率范围内才能够被真正应用（天线的大小）。在传输前需使用调制器将信息信号的频率搬移到信道的频率范围上，传输后在接收器中再使用解调器将信号的频谱搬移回来，如图 1-24 所示。带通传输将在第 3 章介绍。

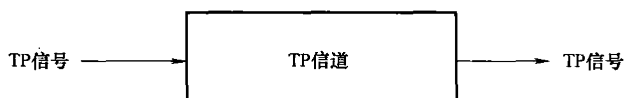


图 1-23 低通信号通过低通信道传输

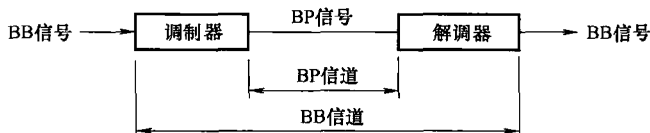


图 1-24 BB 信号的带通传输

图 1-24 表明，带有调制器和解调器的 BP 信道也可以作为 BB 信道来理解。这种观点很有帮助，因为例如在通过无线来进行数据传输时（BP 传输，参见第 3 章），也必须使用 BB 传输（参见第 2 章）的所有方法，这样可以省去复制。图 1-24 也可以看成是图 1-23 的细化，而图 1-23 与图 1-15 具有相同的结构。使用这种方法可以隐藏实际的传输媒介，因为最终信息是通过同轴电缆（TP 信道）来传输的，还是使用微波技术（BP 信道）传输，或是使用光波导进行数字传输对用户来说是无所谓的。

同样使用 A/D 转换器或是 D/A 转换器的数字信道也可以理解成是一个模拟信道，如图 1-25 所示。此外两个图的混合也是可行的，如图 1-26 所示。这样它可以处理例如用无线信道来传输经过数字化的模拟信号（例如 PCM）。

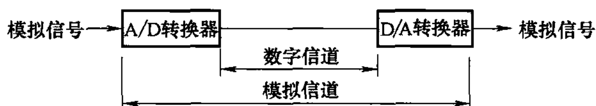


图 1-25 模拟信号通过数字信道传输

信道的界限以及信源和信宿的分界是很自由的。图 1-26 中发送器框之前的所有部分都属于信源。之间的所有单元都可以看成是信道。信道的最内层是传输媒介。这种选择的自由性同时包含对所选择分界线的明确定义的限制。

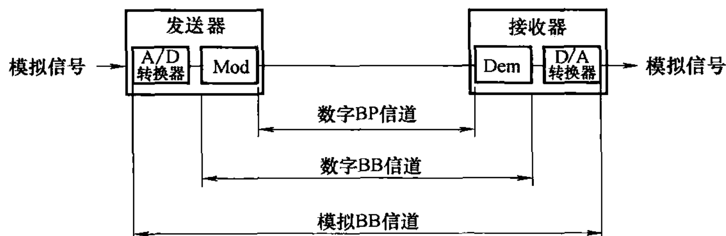


图 1-26 模拟 BB 信号通过数字 BP 信道传输

现在我们要观察一下传输信道的另一个完全不同但很实用的方面：方向的分离。首先一个小小的话题偏离是必要的，例如电话网就是由此而生的。

许多电话用户都希望能临时自由选择通信对象。这个问题使用由无数信道组成，并且根据需要两两相连的网络就能解决。人们称之为交换，由交换机来完成。在两个大城市之间有无数多这样的并行信道，通过复用可以同时使用一种媒介（例如光波导、同轴电缆、无线广播）很便宜地进行传输。

在终端用户范围内（电话机到第一个交换机之间的范围）复用是不可行的，因为连接并不是并行的。每个电话设备使用单独的一条电缆与下一个交换机相连。

绝大多数连接都是所谓的双向连接，即每个方向都可以同时输入或输出。例如电话连接就属于这种情况，通信双方可以同时讲话。大多数数据连接也是双向的，例如为了对数据进行证实（ARQ 系统）或在访问数据库时进行电子对话。

成百万的电话连接导致在终端用户范围内需要有无数的铜线。传统的双向连接每个用户需要四条铜线（四线运行），采用混合器的方向分离法会使复杂性减半。方向分离的首要优点是将费用减半。

方向分离可采用下面的方法：相同位置方法（采用混合电路，只有在有线连接时可行）、分离状态方法（频分复用，FDD）、乒乓方法（时分复用，TDD）。

相同位置方法

图 1-27 所示为四线运行时电话连接的线路，要一个连接使用两条线。两种运行模式之间的过渡采用混合电路或简称混合器（英文为 hybrid）来实现。

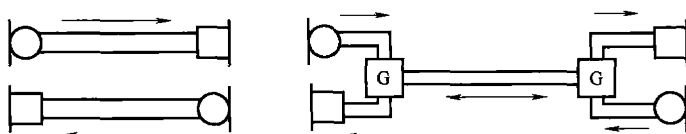


图 1-27 四线运行（左）与两线运行（右）（G 为混合器）

混合电路可以采用有源或无源的方法来实现，后者有 3dB 的衰减。图 1-28 给出了混合电路的组成原理。它由传声器、耳机、双线电阻和一个等价线（双线电阻中的一个相当于一个外界不能接入的阻抗）组成。当所有四个阻抗的值相等时，桥路达到平衡，对角线解耦合。

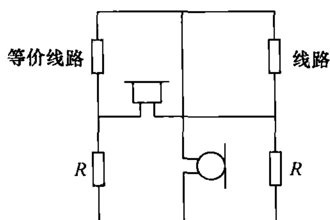


图 1-28 混合电路的组成原理

如果输入阻抗的值发生一点变化，等价线会因此设置一个折中值。这会导致非对称并因此会产生反射。

在用户和本地交换机之间因费用的关系使用的是双线运行。更远的距离（不同交换机之间）使用的是四线运行，如图 1-29 所示。这是因为必要的放大器和复用设备只能是单向运行的（非可逆 2 端子）。如果要像图 1-30 那样在连接路径的每个放大器前后都放置一个混合电路，混合电路的反射所造成的反

向耦合会引起很大的问题。受放大系数的限制需要提高放大器的个数，因此会导致信噪比的降低。

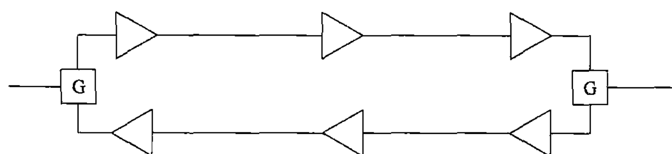


图 1-29 使用中间放大器的四线远程连接

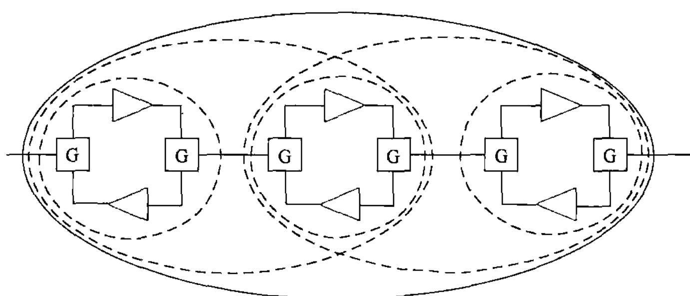


图 1-30 不合适的双线远程连接（圆形和椭圆形虚线表示可能的反向耦合）

四线连接时也存在反向耦合路径，即由于两个终端设备的混合电路所造成的。数据传输时这会导致干扰回声，它们必须要使用特殊的电路来抑制（回声补偿）。此时还要区分近端回声（自己的混合电路造成的）以及远端回声（由通信对端的混合电路引起的）。

分离状态方法

方向由不同的频率范围来区分，并且使用滤波器来进行分离。在传输音频时至少每个方向都要有一个调制解调器。这样传输网络需要处理完全不同的信号，因此该方法没有用在电话网的话音传输中。

这种方法广泛运用于使用电话网进行数据传输所使用的调制解调器中。计算机用两种不同的电平表示“0”和“1”以提供数据。这些信号包含低频成分，不能通过电话系统，针对话音传输而进行优化的 300～3400Hz 这个频率范围。因此调制解调器使用两个不同的频率来代表“0”和“1”。方向的分离很简单，使用两个不同的频率对即可。但所有的 4 个频率都要位于 300～3400Hz 这个频率范围内，这样交换机就意识不到还有其他内容在传输。例如调制解调器标准 V.21（300 波特）规定的频率对为 980Hz/1180Hz 和 1650Hz/1850Hz。

因为方向的分离是根据频率来进行的，所以将该方法称为 Frequency Division Duplexing（频分双工，FDD）。在光波导中也使用这种方法，此时使用不同波长的光来进行处理。FDD 常用于无线通信中（特别是卫星通信）。

乒乓方法

此时方向的分离不是在频域内而是在时域内进行，因此该方法也称作 Time Division Duplexing（时分双工，TDD），即两个站点交替作为发送器和接收器。使用数据缓冲器就可以对外提供连续的数据流。

TDD 在现代无线对讲设备中很常见，但它也用在光纤传输中，此时半透性镜像激光二级管（发送器）和光敏二级管（接收器）连接在光纤的两端。在非对称，即两个方向的数据传输速率不同、数据量很大的双向信道中使用 TDD 具有很大的优势。

但在连接建立时使用 TDD 是很难的，因为首先得进行同步，这得通过例如发送器能够接受使用随机时间来控制它的行为来实现。所以一定得避免两个站点都在不断地发送而无人接收的情况。

在这 3 种方法中，混合电路的相同位置方法需要的带宽最小，因此它被使用得最广泛。

图 1-29 所示的四线连接中放大器之间的距离由电缆的衰减和信息信号的带宽所引起的频率响应来确定。例如 2700 路话音信道采用频分复用（FDM）进行传输时放大器之间的距离为 4.65km。10800 路话音信道时距离下降到 1.55km。远距离传输时在传输路径上很快就需要上百个放大器。因为 FDM 用在模拟传输中，多级信号处理会引起可听见的噪声。相反采用时分复用（TDM）方法的数字传输时，洲际之间的连接与两个村庄之间的连接没有任何质量上区别。

并行线之间会有电导和电容的耦合，这会导致串扰。串扰有如下的类型：可理解的串扰（模拟 BB 传输时）、不可理解的串扰（采用复用的方法进行传输时，外部表现为噪声）、近端串扰（NEXT）、远端串扰（FEXT）。

图 1-31 所示为从上面信道到下面信道的串扰路径，此时每条信道都采用复用的方式工作。当然下面信道到上面信道也有串扰。当电缆上的信号强度相等时对两个方向的影响是相等的。当信号强度不同时，几乎毫无例外地，强度大的电缆会对强度小的电缆产生影响。同轴电缆在高频时受集肤效应的影响不会产生串扰，同样光波导（LWL）相互之间也没有耦合。

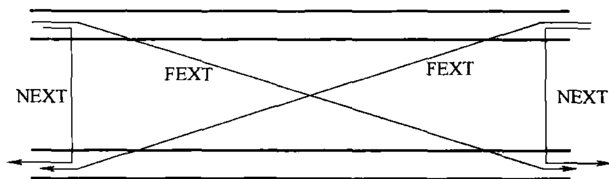


图 1-31 近端和远端串扰

值得一提的是同样需要进行复用的虚连接，如图 1-32 所示。

在传输技术中人们对不同的运行类型进行区别，这里有线连接和无线连接与传统定义的有些不同，因为无线设备不能同时使用相同的频率进行发送和接收。

有线连接时的运行类型：

1）简单运行：单向传输（盲发，没有证实）。

2）半双工运行：同一信道，两个方向交替运行（交替收发信息）。

3）双工运行（全双工）：在两个方向上同时传输。

无线连接时的运行类型：

1）盲发：只是单向传输。例如：无线广播。

2）简单运行：在同一信道上交替传输。

3）半双工运行：每个方向上有一个独立的信道，两个信道交替运行。例如：带有中继的无线连接。

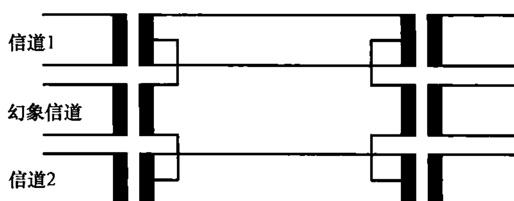


图 1-32 四线电缆中的虚信道

4）双工运行（全双工）：同时使用两个信道。例如：定向广播。

在本节的最后我们还要看一下传输信道的数学模型。它的目的是从数量上估算信道对有用信号的影响，以便在不同调制方法之间进行比较。

绝大多数信道都是线性的，或至少能用足够好的方法进行线性逼近。因此信道可以描述为 LTI（Linear Time Invariant）系统，如图 1-33 所示。数字信道情况下称为 LTD（Linear Time Invariant Discrete）系统。在信号处理相关文献（参见参考文献【Mey00】或【Mey02】）中会详细解释这些系统。

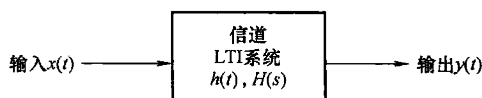


图 1-33 信道作为 LTI 系统

一个 LTI 系统可以完全由它的脉冲响应 $h(t)$ 或传输函数 $H(s)$ 来描述。这个模型可以描述被传输的信号（比较 1.2 节）的线性变化。一种简单的极端情况下传输函数可以退化成只有几个常数，此时能够描述信道与频率无关的衰减情况。

信号在信道中会因为干扰而变形。大多数情况下会叠加一个干扰信号。尽管这些干扰的物理入侵点不能都被定位（通常情况下干扰分布在整个传输线路上），但由于是线性模型可以假设有一个中央入侵点。通常情况下它位于图 1-33 所示的信道之后，因为在这里可以很容易测量干扰的总和。图 1-34 给出了扩展后的模型。

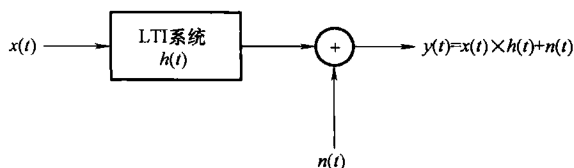


图 1-34 加性干扰的 LTI 信道

图 1-34 中干扰信号 $n(t)$ 通常被描述成高斯噪声，它是由电阻、放大器、大气、银河系等所引起的噪声。这种假设是有意义的，因为根据统计学的中心极限值定理可以由多个随机过程的和计算出高斯过程。

声学水下信道及短波表现出很强的时变特性。这在上面的模型中需要特别关注，此时冲激响应 $h(t)$ 是时变的，但在很短的时间片内保持为常数。

多径信道（无线移动通信、短波广播、电视）具有时延，会在接收端接收到多个不同程度变弱了的原始信号（产生重影）。这种情况下脉冲响应具有下面的形式：

$$h(t) = \sum_{k=1}^{L} a_k \delta(t - \tau_k) \tag{1-15}$$

式（1-15）描述了 L 个衰减为 a_k，时延为 τ_k 的散射路径的情况。但是单个信号并没有出现失真。此时图 1-34 输出端的信号为

$$y(t) = \sum_{k=1}^{L} a_k x(t - \tau_k) + n(t) \tag{1-16}$$

这个模型会覆盖实际中的绝大多数情况。进一步内容可参见参考文献【Gen98】。

1.1.8　信道容量和信息体

一个没有干扰、带宽有限、幅度为有限值的模拟信号所包含的信息量是无限大的，即这个信号可以看成可以无限细微量化的信号。

受有限的控制方法的限制，技术上信号的幅度值始终是有限的。除此之外这些值还不断受或多或少的较大噪声的影响。例如当最大幅度为 10V、所包含的噪声部分幅度为 1V 时，最高可以有 10 个量化等级可以区分。如果噪声幅度变小或信号最大幅度变大，相应地可有无数多的量化等级，因此可以包括较大的信息量。对一个信号的信息量来说，信噪比（Signal to Noise Ratio，S/N）是一个起决定作用的值。信噪比是信号与噪声的比值，多数情况下以 dB 为单位。

与信噪比紧密关联的是动态性。将它理解为最大信号电平（大于这个值就会受到例如放大器的控制极限的限制）与可识别到的最小信号电平（小于这个值就会受到例如放大器的固有噪声的限制）之比。通常动态性的单位也是 dB，它表示噪声部分的百分比。

所有的信息都是量化的，也能使用数字信号（时间和幅度都是离散的）来描述。模拟信号情况下可以假定量化级别间隔可以通过噪声的幅度来给出。信号改变量化级别所需时间与最大改变速度有关，因此与带宽也是有关的。最大改变速度与带宽之间的关系可以根据傅里叶变换的特性，即时间带宽积给出。

因此信号的信息量与下面两点有关：

1）可区别的量化等级的数量，即信噪比及信号动态性。

2）信号最大改变速度，即它的带宽 B。

物理信号的带宽、幅度是有限的，并且含有一定比例的噪声。所以信噪比永远不能小至可以忽略，并且因此信息量总是有限的，这是因为所有信息最终都是数字化的。如果信道进一步地限制了带宽，并且又给信号增加了额外的噪声，信息量又会减少。

香农在 1948 年给出了信道容量的概念。他的假设前提是接收端信号会叠加一个高斯噪声。信道容量 C 是信道能够传输的用 bit/s 来表示的最大的信息传输速率。将式（1-1）除以 T 可以得到

$$C = B\log_2\left(P_S + \frac{P_N}{P_N}\right) = B\log_2\left(1 + \frac{P_S}{P_N}\right) \tag{1-17}$$

式中　C——用 bit/s 表示的信道容量；

B——用 Hz 表示的信道带宽；

P_S——信号功率；

P_N——噪声功率。

正如所期望的那样在式（1-17）中，带宽和信号功率与噪声功率的比（即信噪比）是变量。人们将表达式 $\log_2\left(1+\frac{P_S}{P_N}\right)$称为信道动态性。

换算成以十为底的对数有

$$\log_2(x)=\frac{\lg(x)}{\lg(2)}\approx\frac{\lg(x)}{0.3}=\frac{\lg(x)}{3}\cdot 10$$

从现在开始我们用 lg(x) 表示以十为底的对数。根据式（1-17）有

$$C\approx\frac{B}{3}10\lg\left(1+\frac{P_S}{P_N}\right)$$

通常情况下 $P_S \gg P_N$，因此上式可以简化为

$$C\approx\frac{B}{3}10\lg\left(\frac{P_S}{P_N}\right)=\frac{B}{3}SR \tag{1-18}$$

其中 C 的单位为 bit/s，B 的单位为 Hz，SR 的单位为 dB，上式中 SR 是用 dB 来表示的信噪比。表 1-7 是不同信道容量的概况，其中的信噪比是与应用相对应的值（归一化值）：

表 1-7 模拟信道信道容量举例（归一化的信噪比）

业务	B/kHz	(P_S/P_N)/dB	C/(kbit/s)
电传打字机	0.025	15	0.13
电话	3.1	40	41
UKW 无线广播	15	60	300
CD(模拟输出)	20	96	640
电视	5000	45	75000

CD 情况下量化噪声决定了信噪比。A/D 转换器中字宽要增加一个 Bit 能将量化电压减半、量化噪声功率减小为原来的 1/4，但实际上并没减少信号功率。因此每 Bit 量化信噪比增加了 6dB。使用 16Bit 字宽信噪比约为 96dB。前面数据传输速率按 16 × 44100（采样频率）≈700kBit/s 计算，因此会与表中所提到的 640000bit/s 有区别，因为根据采样定理只允许用 22kHz 代替 20kHz 来计算。

根据表 1-7 及上面的考虑可以推导出不同信号的 PCM 字宽：

1）话音（Voice）：8Bit；

2）视频（Video）：8～10Bit；

3）音频（Audio）：14～16Bit。

这点已经成功的信息传输信噪比必须足够大而不是信号的功率要足够大，式（1-17）也反映了这点。从表 1-8 可以找到一点话音传输（即图 1-2 中的信宿是人的耳朵）所能容忍的最小信噪比的感觉。

表 1-8 话音传输时的最小信噪比

应 用	最小信噪比/dB	应 用	最小信噪比/dB
电话	33	接入电话网的无线广播	15
MW 无线广播	33	熟练话务员,母语	6

无线话音连接（例如外海上的船只）只有在信噪比大于15dB 时才能进入到电话网中，否则可理解性太差。熟练的话务员在很差的情况下也能完成连接，当然这里在大部分内容是使用母语的情况下。此时可以充分利用语言的冗余性。当检测可理解性时不希望使用前面曾提到过的含糊音。

这种效果人们在日常生活中也能观察到。只懂英语的人能听懂新闻广播，但并不一定能听懂流行歌手，因为后者的话音信号因伴奏音乐的影响会产生一个较小的信噪比。相比之下德语流行音乐的歌词比较容易理解。

将式（1-17）乘以传输时间 T，就可以得到在时间 T 内，容量为 C 的信道所能传输的息量总量：

$$I = C \cdot T = B\log_2\left(1 + \frac{P_S}{P_N}\right) \cdot T = B \cdot D \cdot T \tag{1-19}$$

式中 D——前面引入的信道动态性，它描述了对干扰的敏感性。

因此除了这里指信道的动态性，而式（1-1）是指信号的动态性之外，式（1-19）同式（1-1）是完全相同的。

根据式（1-19）可以将传输的信息总量 I 以及表示这些信息的信号理解成边长分别为 B、D（信号动态性）和 T 的立方体的体积。人们称之为信息立方体、消息立方体或香农立方体。信道本身的特性可以根据式（1-17）用信道窗口，即边长为 B 和 D（信道动态性）的长方形来刻画。现在信号适配（调制、信号编码）的任务是对用变量 B、D 和 T 表示的消息立方体进行变形，在保持立方体体积不变的前提下使新立方体能够适应信道窗口，如图 1-35 所示。

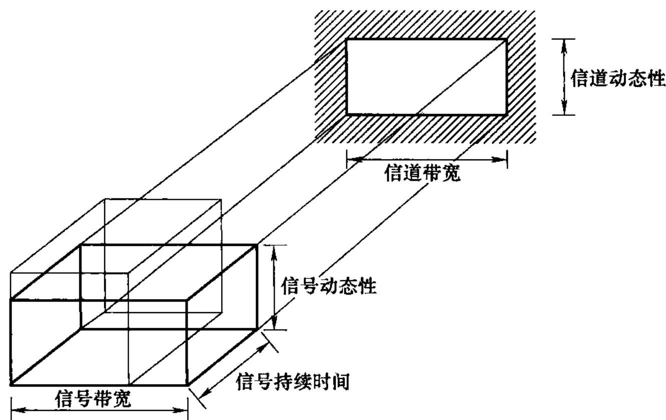


图 1-35 消息立方体（位于前面，用细线表示）、适配后的立方体（位于前面，用粗线表示）和信道窗口（位于后面，用粗线表示）

如果信号的带宽大于信道带宽，那么就会出现信息丢失。同样，当信号的动态性大于信道的动态性时也会出现这种情况。反之信道就没有得到充分的利用。当出现如图 1-35 所示的情况时，信号动态性压缩得要比信号带宽扩展得小，这样相应的传输时间才会减少。

如果多个相互无关的信息（消息信号）要同时经过相同的信道传输（复用或集群传输），那么可以将这些信息适配在一个公共的信息立方体中，这个公共立方体与信道窗口是适配的。根据汇集方式的不同，可以称之为时分复用、频分复用、幅度复用（很少使用）以及功能或码分复用等，如图 1-36 所示。最后一种方法是很现代的，在扩频传输（带宽扩展技术，参见 5.3 节）中会用到。

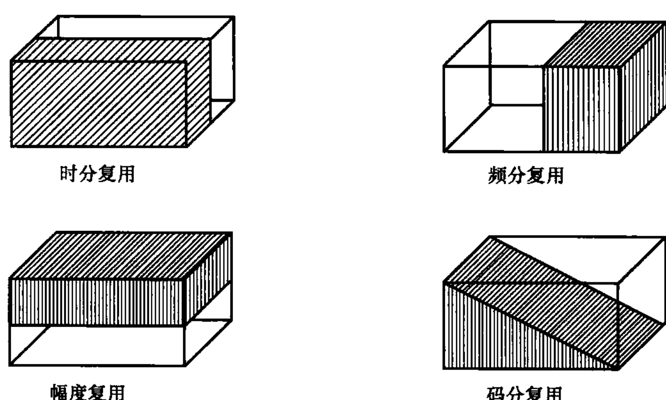


图 1-36 两种（不一定非得具有相同大小的）信息量的合并

受信道干扰的影响在传输中会有部分信息丢失（含糊信息），也会有一些新的伪造的信息加入进来（不相关，垃圾信息熵）。而原始信息中真正到达接收端的被称为传递信息，如图 1-37 所示。式（1-17）中所给出的信道容量涉及到的就是这部分信息。

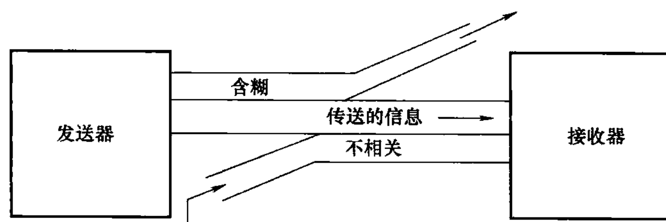


图 1-37 信道中的干扰所引起的信息流的变化

复用指的是多个小信息流合并成一个较大的信息流。反过来，即将一个较大的信息流分成多个小信息流称为向下复用，它的目的是可以使用多个小容量的信道并行传输，如图 1-38 所示。

信号动态性和信道动态性需要进行进一步的解释。为此我们从比较容易理解的信道的动态性开始。

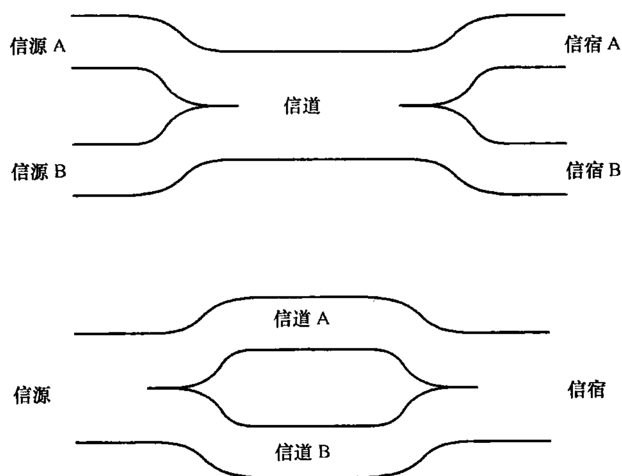


图 1-38　向上复用（上）和向下复用（下）

一个可实施的放大器电路有一个可控制的范围，例如 ±10V。如果超过这个幅度信号就会因为过饱和而产生一个非线性失真。除此之外，放大器还有一个固有噪声，瞬时值应该在 ±1mV 之间。放大器输出端的信噪比与有用信号的有效功率和噪声的有效功率之比有关，即与电压的有效值之比有关。当有用信号的最大值为 10V 时，才可以根据信号形状来确定它的有效功率。

如果有效信号包含有信息，那么它不能是确定的。因此我们假设不管是有用信号还是噪声信号都是在 ±A 范围内等分布的随机信号，$A_S=10V$，$A_N=1mV$。这种情况下功率为 $A^2/3$（计算在 3.3.1 节中给出）。因此所描述的放大器的信噪比为

$$\frac{P_S}{P_N}=\frac{(10V)^2/3}{(1mV)^2/3}=\frac{10^2}{10^{-6}}=10^8 \hat{=} 80dB$$

从工艺上来说每种信道都有一个可控范围，因此 P_S 是受限的。同样每种信道由于工艺也有它的固有噪声，各种类型的陌生噪声也都计算在内。使用这种方法信道输出端的信噪比也被迫受限。利用这个信噪比根据式（1-19）可以构造出信道动态性的单一图。对于定性思考，使用信道动态性还是输出端的信噪比是无所谓的。通常将用 dB 表示的信噪比作为动态性的描述方法。从定性的角度来看，动态性和信道动态性是相同的，而数量上则是不同的。

现在我们看看信号的动态性！它同样与信噪比有关，问题是只讨论有用信号不能区分出信号和噪声。

正如前面已经解释过的，所有的信息最终都是要进行数字化的，即是可数的。相反所有的物理信号最终都是模拟的，即是不可数的。用无限多可能幅度值中的有限部分值来描述一定的含义，信号就是通过这种方法来携带信息的。这个部分值的范围就是决定了信号的动态性。部分值越多，那么信号变化例如 1mV 就变成其他含义。如果部分值较少，那么变化 1mV 毫无意义。

信号动态性与信道动态性相反，它是主观的大小值。例如在话音信号中较小的背景噪声对话音的理解影响不大，因此可以说话音信号需要的信噪比为 40dB（见表 1-7）。在高质量的音乐中人们对很小的细节都很感兴趣，因此要求具有从 60～90dB 的较大的动态性。

当信噪比为 100dB 时，1V 的信号幅度所要求的噪声幅度为 10μV。如果这个信号用理想放大器，即没有噪声的放大器放大 10 倍，信号的幅度为 10V，噪声的幅度变成 100μV，信噪比仍然为 100dB。但若采用上面可实施的放大器，信号的幅度仍然为 10V，但由于放大器的固有噪声，噪声信号的幅度变成 1mV，因此信噪比只有 80dB。

如果信号只有 40dB 的信噪比（信号幅度为 1V，噪声幅度为 10mV），那么放大后，信号的幅度为 10V，被放大了的噪声的幅度为 100mV，而放大器的固有噪声为 1mV。那么信噪比仍然约为 40dB。

每个源信号都是“纯的”，即没有噪声的。但连传声器都能给音乐信号加上一些噪声。每一个后续的处理（存储、传输等）都进一步降低了信噪比。在信道的末端信号的动态性最高能等于信道的动态性。因此信号的动态性要与信道的动态性相匹配，可能要与信道编码或调制相匹配，如图 1-35 所示。

正如前面已经提到的，信号的动态性是主观的，即它是由应用来确定的。表 1-7 和表 1-8 给出了所要遵守的归一化以及最小信噪比的要求。

> 信号动态性是一个主观的大小值，描述所感兴趣的幅度受限的信号的细节程度。
>
> 信道动态性是一个客观的大小值，描述在信道输出端所能获得的、具有一定细节程度的信道的能力。
>
> 具有较大动态性的信号要求信道也具有较大的动态性，即较大的信噪比。因此信号才具有较强的抗干扰性。

式（1-17）与信号类型有着很有意思的关系。如图 1-39 所示，根据时间和数值轴的连续和离散两个标准可以将信号分成 4 类。

图 1-39 的信号 b）和 d）通常不被描述成脉冲位图，而是采样值序列（加权的 Dirac冲击）。但信息量还是相同的，比较图 1-9。

根据式（1-17），考虑到信息量，对 4 类信号的解释如下：

1）采样后信号的带宽一定是受限的，即 $B<\infty$。

2）对于幅度连续的信号，可以将它的值设为 π 或 e，即是个无理数。这在用十进制编码或二进制编码的情况下只能用无限多位数才能准确表示。但在量化过程中用有限多的位数来表示就足够了。

理论上图 1-39 的信号 a）和 b）包含有无数多的信息，因为为了表示它的瞬时值需要无数多的位数。根据式（1-2）正确采样的信号的带宽是有限的，可用有限个数的采样值来准确地描述。信号 a）和 c）（是时间连续的）理论上带宽也是无限的。

因此只有数字信号才能用有限个数的采样值和有限的位数进行准确地描述。数字

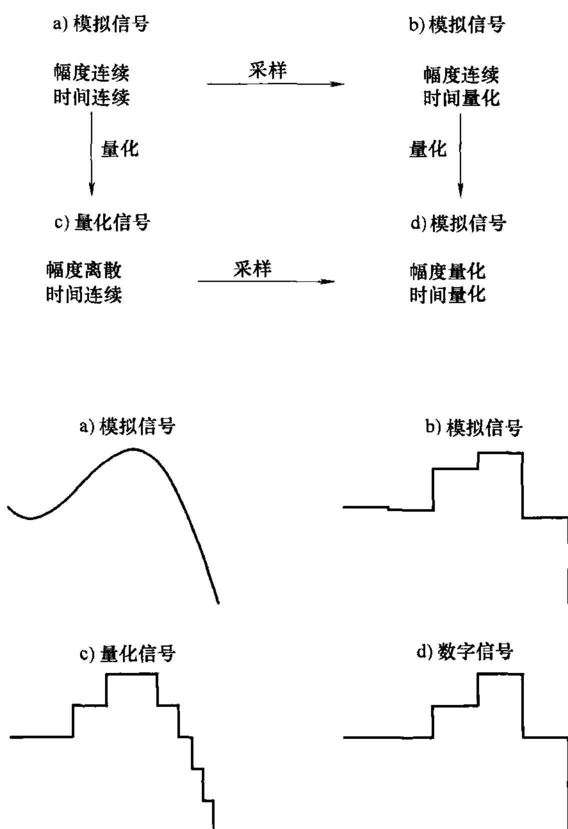


图 1-39 信号被分成 4 类

信号根本不能携带无限多的信息，因此是为了信息（这些信息最终也是可数的）通过一个技术系统（它最终只能无误差地传输有限多的信息）进行传输而定制的。所以只有数字信号才能被无误差地再生，因此数字信号在技术上是很重要的。

现在利用信息论的一点知识可以推导出理论上的通信极限。根据式（1-17）可以采用两种方式增大信道容量：

1）通过增加信噪比→C 可以任意增加；

2）通过增大带宽→C 可以增大到有限的极限。

通过增大 B 也可以增加 P_N，因此 C 还是有限的。热噪声是由电子在导体或半导体中的不规则运动而引起的。它随着温度的增高而加大并且因有无数的电子参加而具有高斯分布。功率密度频谱一直到 10^{12} Hz（即在整个技术上所感兴趣的范围）都是与频率无关的（即白噪声），值为 N_0。受这种类型干扰的信道被称为高斯信道或 AWGN（Additive White Gaussian Noise）信道。

热噪声功率：

$$P_N = k \cdot T \cdot B = N_0 \cdot B \tag{1-20}$$

式中 k——Boltzmann 常数，$k=1.38\cdot10^{-23}$J/K；

T——绝对温度，单位为 K；

B——带宽，单位为 Hz；

N_0——单位频谱热噪声功率密度，单位为 W/Hz 或 Ws。

将式（1-20）代入到式（1-17）中有

$$C=B\log_2\left(1+\frac{P_S}{kTB}\right)=B\log_2\left(1+\frac{P_S}{N_0B}\right) \tag{1-21}$$

如果增大 B，那么括号内第二部分的值就会减小。它的极值可以根据自然对数的展开式来计算：

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\cdots\approx x \qquad (当\ x\ 很小时)$$

因此以二为底的对数有

$$\log_2(1+x)=\frac{\ln(1+x)}{\ln(2)}\approx\frac{x}{\ln(2)} \tag{1-22}$$

所以当 $B\to\infty$ 时最大信道容量为

$$C_{\max}=\frac{B}{\ln(2)}\cdot\frac{P_S}{kTB} \qquad C_{\max}=\frac{P_S}{k\cdot T\cdot\ln(2)}=\frac{P_S}{N_0\cdot\ln(2)} \tag{1-23}$$

商 P_S/C 表示每 bit/s 的功率。它与功率乘以 s/bit 是等同的，即每个 bit 的能量 E_{bit}。根据式（1-23）计算 $P_S/C_{\max}$，可以获得为了在无限宽的信道中传输 1bit 而在接收器输入端所需要的最小能量。在室温（$T=293$K）中这个值为

$$E_{\text{bit}_{\min}}=kT\ln(2)=2.8\times10^{-21}\text{Ws} \tag{1-24}$$

这样信息与能量就相互关联上了。专业的相对论将质量和能量联系在一起，因此这 3 个值都是相关的。这在本书的第一节中就已经说过了。式（1-23）的另一种计算途径是使用量子物理学的知识，得出的结果是相同的。能量对信息传输来说是必要的，这点从式（1-17）就已经能够看出来。如果令式（1-17）中的 $P_S=0$，那么括号中第二项的值就会消失。1 的对数与底无关都为 0，因此 $C=0$。

根据式（1-19）信息量 I 的传输需要带宽、功率、时间和金钱。关于最后一点信息论是不讨论的。同样对可能的较大的时延讨论得也不多，例如在引入信源和信道编码时可能遇到这种情况。

通常情况下用户对电信系统所能传输的用 bit 表示的信息量并不太感兴趣，而对每秒钟所能传输的信息量，即用 bit/s 表示的信息速率 J 比较感兴趣。J 是描述连接有效性的比较流行的方法，最大可以等于信道容量 C。重要的还有效率，即单位有效性。J 与下面的内容有关：

1）投入的功率→每 bit 的能量；

2）投入的带宽→单位频谱的信息速率；

3）投入的金钱（系统花费、计算开销、需要的存储）。

通常情况下并不计算 J/P_S，而是计算它的倒数 P_S/J，即每 bit/s 的功率。这正是上面所引入的每 bit 的能量：

$$E_{bit} = \frac{P_S}{J} \geqslant \frac{P_S}{C}\left[\frac{Ws}{bit}\right] \tag{1-25}$$

此外人们通常也将 E_{bit}对热噪声功率密度进行归一化。这个比值与温度有关，并且没有量纲（[bit] 是个没有量纲的数），因此也能用 dB 来描述。对极限值 C 有

$$\frac{E_{bit}}{N_0} = \frac{P_S}{CN_0} = \frac{P_S}{CkT}\left[\frac{Ws}{bit \cdot Ws}\right] = [1] \tag{1-26}$$

这样式（1-17）可以换一种方式来表示：

$$C = B\log_2\left(1 + \frac{P_S}{P_N}\right) = B\log_2\left(1 + \frac{E_{bit}C}{P_N}\right) = B\log_2\left(1 + \frac{E_{bit}C}{N_0 B}\right)$$

现在也可以对带宽进行归一化：

$$\frac{J}{B} \leqslant \frac{C}{B} = \log_2\left(1 + \frac{C}{B} \cdot \frac{E_{bit}}{N_0}\right) \tag{1-27}$$

图 1-40 用对数形式给出了式（1-27）的值。在曲线之下（不等号）无差错通信是可行的，曲线之上通信是不可靠的。位于线（等号）上的系统能达到理论上可能的信道容量，它从信息论的角度上来看是最优的，但技术上过于复杂，只有用于研究宇宙的太空飞船工作在接近这个极限的值上。

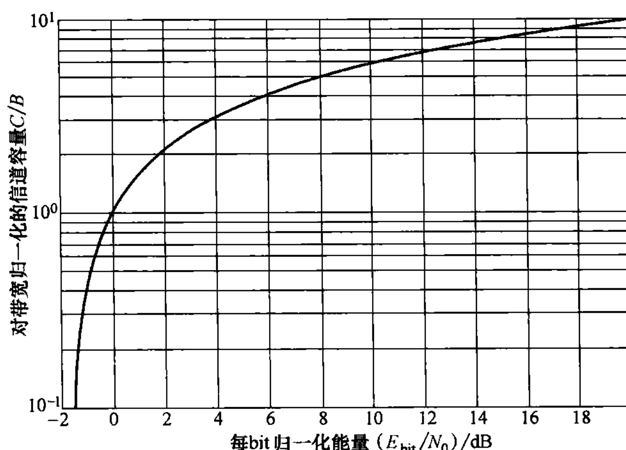


图 1-40　最大单位频谱信息传输速率（即对带宽归一化的信道容量）与对 N_0 归一化的每 bit 能量的函数关系

图 1-40 中每 bit 能量及信号功率从左到右是不断增长的。作为补偿，带宽一定是下降的。因此左下方是使用无限的带宽，但尽可能小的功率来传输。极限值可以根据式（1-28）来计算，此时当 $B \to \infty$ 时，log 函数的参数会变得很小，可以利用式（1-22）的简化公式。于是有

$$\frac{C}{B} \leqslant \frac{C}{B} \cdot \frac{E_{bit}}{N_0} \cdot \frac{1}{\ln(2)} \quad \frac{E_{bit}}{N_0} \geqslant \ln(2) = 0.693 \hat{=} -1.6\text{dB} \tag{1-28}$$

这是通过高斯信道进行无差错传输的信噪比的绝对下限，被称为香农极限。当

E_{bit}/N_0的值较大时可以进行无差错传输。但是为此需要一个最优（可能是无限复杂的）的信道编码。

在室温下传输 1bit 所需要的最小能量可以根据式（1-28）来计算，这与式（1-24）是一致的：

$$E_{\text{bit}_{\min}} \geqslant N_0 \ln(2) = kT\ln(2) = 2.8 \times 10^{-21}\text{Ws}$$

当增大带宽时，C 会逐渐增大到一个极限值。但增大信号功率时，C 会任意地增大。另外很有趣的一点是用对功率进行归一化的信道容量来代替对带宽进行归一化的信道容量。此外人们还将 P_S用 N_0来归一化，这样又可以得到一个无量纲的值：

$$C(\text{bit/s}) \rightarrow \frac{C}{P_S/N_0} = \frac{CN_0}{P_S}\left[\frac{(\text{bit/s})\cdot \text{Ws}}{\text{W}}\right] = [1]$$

根据式（1-21）又可以得到

$$\frac{C}{P_S/N_0} = \frac{B}{P_S/N_0}\log_2\left(1 + \frac{1}{\dfrac{B}{P_S/N_0}}\right) \tag{1-29}$$

图 1-41 给出了这个公式的值。

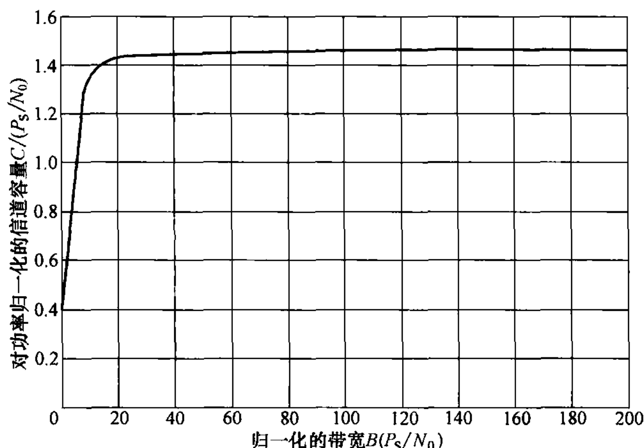


图 1-41　信道容量与带宽的函数（对功率归一化）

从图 1-41 中可以看到一个极限值 1.44，这个值可以根据式（1-23）计算出：

$$\frac{C}{P_S/N_0} \leqslant \frac{1}{\ln(2)} = 1.44 \tag{1-30}$$

图 1-40 的左下方满足式（1-28）： $\dfrac{E_{bit}}{N_0} \geqslant \ln(2)$

将式（1-25）代入，并且求解 $C/(P_S/N_0)$，于是又能得出式（1-30）：

$$\frac{E_{bit}}{N_0} = \frac{\dfrac{P_S}{C}}{N_0} = \frac{P_S/N_0}{C} \geqslant \ln(2) \rightarrow \frac{C}{P_S/N_0} \leqslant \frac{1}{\ln(2)} = 1.44$$

1bit 并不等于 1Bit。这里也必需准确地区别下面各个符号之间的关系：

J 是用 bit/s 表示的信息速率；

C 是用 bit/s 表示的信道容量；

R 是用 Bit/s 表示的数据速率。

香农已经证明，$J \leqslant C$ 必需永远得到满足。

因为 1Bit 最多可以表示 1bit，因此还会有：$R \geqslant J$

$R = J$ 意味着数据流是没有冗余的。

$R > J$ 可以通过信道编码来实现，即增加冗余。

$R > C$ 是可能的，对有限传输系统甚至是有意义的。

$R < C$ 还意味着 $J < C$，信道没有被充分利用。

$R = C = J$ 从计算上看是可能的，但物理上是不可能的。香农已经计算出了高斯信道的信道容量 C。噪声形式的干扰可能有很高的瞬时值，必然（虽然可能只是很少出现）会引起误码。只有引入信道编码才能消除由不可避免会出现的误码所引起的损害。

> 无信道编码的传输系统的工作状态与理论上可能的极限会相差很远。

现在必需要仔细解释一下带宽这个概念。式（1-18）给人的印象是信息信号有明确定义的带宽。但事实却不是这样，因为从傅里叶变换中可以看出，时间上有限的信号，它的带宽是无限的，反过来也如此。由技术方法产生的信号在时间上总是有限的，但幸运的是它们的频率总是下降得很快。理论上带宽为无限大的频谱，在实际中都可以假设成带宽是有限的，但这个极限需根据不同的方法来得到。

图 1-42 所示为在数据传输中经常会出现的带通信号的功率密度频谱。频谱对称地分布在一个中间频率 f_{Tr}（Tr 为载波，即 Carrier）的两边。幅度密度频谱的包络都遵循 $\sin(x)/x$ 曲线。图 1-42 所示的功率密度频谱也有一个包络，它是按 $(\sin(x)/x)^2$ 曲线下降的。图 1-42 所示的曲线有多种带宽定义方式：

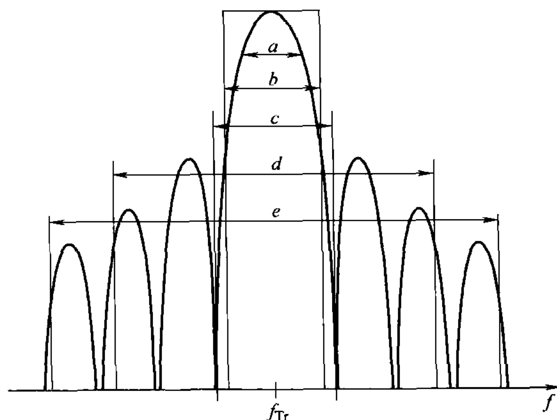


图 1-42　以一个数据信号（解释在正文中）频谱（单位为 dB）为例的不同的带宽定义

a）3dB 带宽：功率密度 $p(f)$ 下降到最大值 $p(f_{\mathrm{Tr}})$ 的一半。

b）等价噪声带宽或等价矩形带宽：被研究的信号的带宽很大，但总功率 P 是有限的。强制地将这些功率分布在一个高度为 $p(f_{\mathrm{Tr}})$ 的矩形频谱中（如图 1-42 中用细线所描述的矩形）。这个矩形的宽度就是等价噪声带宽。

c）主叶宽：频谱包络中最内部两个零点之间的宽度。

d）99%带宽：99%的信号功率都落在这个带宽范围内。

e）极限频谱功率密度：在极限值之外，功率密度与最大值 $p(f_{Tr})$ 相比，至少要下降一个事先定义好的系数倍。通常使用的是 -23dB 极限、 -35dB 极限和 -50dB 极限。

如果它涉及的是估计信道中最大可忍受的带宽极限，这样接收到的数据还需进行解码，那么就可以使用例如带宽定义 c)。如果需要观察的是邻居信道的干扰，那么就会使用 d）或 e)。无论如何所选择的带宽定义必需是明确的。

带宽总是以单边频谱定义，单位是 Hz（频率 f 而不是角频率 ω)。

总结：设计一个通信系统的目标是：最大化信息速率、最小化误码率、最小化发送功率、最小化带宽、最小化系统的复杂性（计算开销、花费)、最大化可用性和可靠性。

很明显，部分单个目标是相互矛盾的。实际上是找一个最优的折中，需要基本条件是：

1）理论物理上的极限（例如奈奎斯特带宽（参见 2.5 节)、香农极限)。

2）技术上的极限（计算能力、存储需求、花费)；

3）规定和管制（例如频谱的使用)；

4）标准；

5）其他基本条件（如卫星运行轨道)。

1.1.9 调制技术简介

调制是指载波信号的一个或多个参数随着载有信息的信号（信息信号）进行变化的过程。通过这个过程载波就加载了信息的内容，但这个过程应该尽可能不改变信息内容。解调在接收端进行，它是调制的逆过程，即从接收到的信号中恢复出信息。图 1-43 给出了一个简单的结构图。

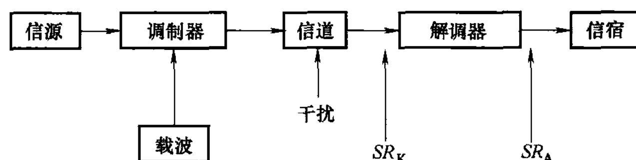


图 1-43 调制和解调过程的简单框图

调制之后信息以另一种方式出现，例如，在另一个（通常是较高的）频率范围内。

如果将调制参数变成时间上和数值上都是离散的，就是数字调制，否则就是模拟调制。

调制的目的有多种（比较 1.1.2 节)：

1）信道适配（物理上的原因)，它又包括频率范围的适配和提高抗干扰能力。

2）复用后进行传输（充分利用信道的容量)（社会生态上的原因)。

原则上任何信号类型（包括噪声）都可以作为载波。但技术实现上只有两种信号类型：

1）谐波载波，它可被用作带通传输和频分复用。

2）脉冲载波，被用在基带传输和时分复用。

信息信号可以是模拟的或数字的。因此调制类型共可以分成 4 类，这些将在本书的 3.1～3.4 节中进行介绍。表 1-9 总结了分类的内容。

调制过程的主要特性就是在已调信号中包含新的频率，因此调制和解调都是一个非线性过程。尽管如此，调制信号和载波信号中被影响到的参数之间的关系可以是线性的。

表 1-9 调制类型一览表

	脉冲载波 （BB 信道）	谐波载波 （BP 信道）
模拟信息信号	3.2 节 PAM（脉冲幅度调制） PFM（脉冲频率调制） PPM（脉冲相位调制） PWM（脉冲宽度调制）	3.1 节 AM（幅度调制） WM（相位调制） SSB（单边带调制） QAM（正交 AM）（模拟）
数字信息信号	3.3 节 PCM（脉冲编码调制） DPCM（差分 PCM） DM（Delta 调制） ADM（可适配 DM）	3.4 节 ASK（幅移键控） FSK（频移键控） PSK（相移键控） QAM（正交 AM）（数字）
	在时间轴上 界限相对严格 →时分复用	在频率轴上 界限相对严格 →频分复用

图 1-44 给出了几种模拟和数字调制的例子。比较模拟频率调制（图 1-44 左下）和数字频率调制（FSK，图 1-9 下）可以得到一个很重要的结论：模拟调制方法和数字调制方法的区别不在于它的本质特点，而在于信息信号的形式。数字调制理论是模拟调制理论发展的一种特殊情况。真正的解码，即对接收到的信号按照后面方便使用的符号所进行的排序（数字信号仅是对模拟信号用数字进行的翻译！），会使用独立的阶段来实现，比较图 1-24。

在图 1-43 中插入了信噪比 SR_K 和 SR_A。SR_K是解调器输入端口的信噪比，用 dB 表示。在理想前端接收的情况下，它的值与信道末端的信噪比是相同的。SR_A是解调器输出端口的信噪比，用 dB 描述，在接收器的后续处理（如放大器）是理想的情况下，它的值就是整个接收器（如音箱）输出端的信噪比。

调制器输出端的信噪比（在图 1-43 中没有给出）在后续处理（如功率放大）理想的情况下就等于发送器的输出端以及信道的输入端的信噪比。这个信噪比的值很大。但在传输路径上信号会衰减并且受干扰的影响，因此在信道的末端，由于受信道的有限动态性的影响信噪比的值会减为 SR_K。

调制能够抑制干扰，即 SR_A 可能大于 SR_K。这是可能的，因为通过调制，可以对信息立方体进行变换，用带宽来交换信噪比，如图 1-35 所示。通过解调来该提高信噪比被称为调制增益 G_M，它用 dB 来表示。

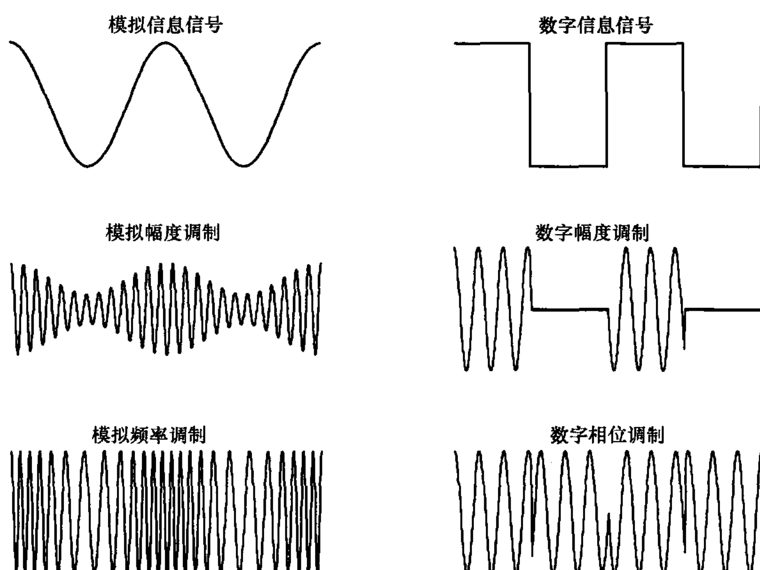


图 1-44 模拟和数字调制类型的例子

$$G_{\mathrm{M}} = SR_{\mathrm{A}} - SR_{\mathrm{K}}(\mathrm{dB}) \tag{1-31}$$

后面会详细描述这种机制。为此要比较解调器输入端口和输出端口的信号。每个信号都有一定的持续时间、带宽和动态性，根据式（1-19）解调器输入端口信号的信息量为：

$$I_{\mathrm{K}} = C_{\mathrm{K}} T_{\mathrm{K}} = B_{\mathrm{K}} \log_2\left(1 + \frac{P_{\mathrm{S_K}}}{P_{\mathrm{N_K}}}\right) \cdot T_{\mathrm{K}}$$

理想解调器不会丢弃信号中的信息，也不会给信号增加新的信息。它只是对立方体进行变形。即 $I_{\mathrm{K}} = I_{\mathrm{A}}$ 是成立的。我们限制在对时间轴不进行改动（这是通常的情况）的调制类型上，即 $T_{\mathrm{K}} = T_{\mathrm{A}}$ 也是成立的。因此会有

$$B_{\mathrm{K}} \log_2\left(1 + \frac{P_{\mathrm{S_K}}}{P_{\mathrm{N_K}}}\right) = B_{\mathrm{A}} \log_2\left(1 + \frac{P_{\mathrm{S_A}}}{P_{\mathrm{N_A}}}\right)$$

$$\frac{B_{\mathrm{K}}}{B_{\mathrm{A}}} \log_2\left(1 + \frac{P_{\mathrm{S_K}}}{P_{\mathrm{N_K}}}\right) = \log_2\left(1 + \frac{P_{\mathrm{S_A}}}{P_{\mathrm{N_A}}}\right)$$

现在令 $\beta = \frac{B_{\mathrm{K}}}{B_{\mathrm{A}}}$，带宽放大可以表示为

$$\beta\log_2\left(1+\frac{P_{S_K}}{P_{N_K}}\right)=\log_2\left(1+\frac{P_{S_A}}{P_{N_A}}\right)$$

$$\log_2\left(\left(1+\frac{P_{S_K}}{P_{N_K}}\right)^\beta\right)=\log_2\left(1+\frac{P_{S_A}}{P_{N_A}}\right) \tag{1-32}$$

$$\left(1+\frac{P_{S_K}}{P_{N_K}}\right)^\beta=1+\frac{P_{S_A}}{P_{N_A}}$$

$$\frac{P_{S_A}}{P_{N_A}}=\left(1+\frac{P_{S_K}}{P_{N_K}}\right)^\beta-1$$

比较公平的是在相同的发送功率 P_S 的情况下对不同的调制方法进行比较。根据式(1-20)，噪声功率 P_N 随着带宽成比例增长。当带宽增大两倍而发送功率保持相同时，信噪比会变成原来的一半。因此当带宽放大系数为 β 时，有

$$\frac{P_{S_A}}{P_{N_A}}=\left(1+\frac{1}{\beta}\cdot\frac{P_{S_K}}{P_{N_K}}\right)^\beta-1$$

现在我们用 dB 来计算功率并将用 dB 引入 SR_K：

$$10\lg\left(1+\frac{P_{S_A}}{P_{N_A}}\right)=10\lg\left(\left(1+\frac{1}{\beta}\cdot\frac{P_{S_K}}{P_{N_K}}\right)^\beta-1\right)$$

$$SR_A=10\lg\left(\left(1+\frac{1}{\beta}\cdot 10^{0.1\cdot SR_K}\right)^\beta-1\right) \tag{1-33}$$

这个公式描述的就是带宽放大系数为 β 时的理想调制方法的调制增益。图 1-45 给出了当 $\beta=1$、$\beta=2$ 和 $\beta=10$ 时的调制增益的曲线。

当 SR_K 较大时可以对式（1-33）进行简化：

$$SR_A\approx 10\lg\left(\frac{1}{B}\cdot 10^{0.1\cdot SR_K}\right)^\beta=\beta 10\lg\left(\frac{1}{\beta}\right)+\beta\cdot SR_K \tag{1-34}$$

这个公式描述了可能的最好情况：当带宽放大系数 β 线性增长时，解调器后的信噪比是呈指数增长的。绝大多数调制方法都没有这么好，只有 PCM 的 SR_A 表现出是呈指数增长的。

根据图 1-45，在第 3 章中将对实际的调制方法进行测量。$\beta=1$ 意味着通过调制带宽没有发生变化，信息立方体没有进行变形而只是在频率轴上发生了平移，单边带调制（SSB）（幅度调制的一种特殊方式）就是这种情况。当然 SSB 不会获得任何调制增益（即 $SR_A=SR_K$），因此只是作为参考方法。任何曲线在 SSB 曲线之上的调制方法都能产生调制增益，频率调制和已经说过的 PCM 就属于这些方法。在长波、中波和短波无线广播中使用的传统 AM 甚至还有损失，它们的曲线在 SSB 的曲线之下。调制增益 G_M 是图 1-45 中与 $\beta=1$ 的曲线的差，这从图中可以得到。

调制增益，例如 10dB，意味着信号的发送功率与 SSB 发送器相比可以减少 10dB，但在解调器输出端两种信号的质量却是相同的，即有相同的信噪比。

在数字传输中不是信噪比而是误码率（Bit Error Rate，BER）起着非常关键的作用。随着信噪比的下降 BER 会逐渐增大。在 PCM 情况下，在 D/A 转换器之后信噪比

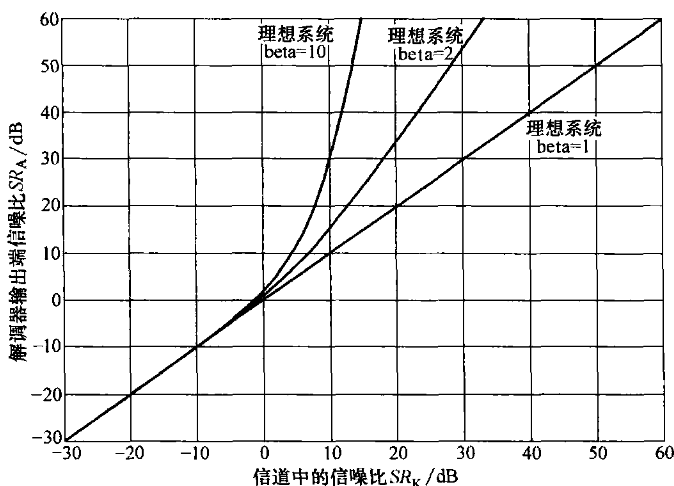


图 1-45 不同带宽放大系数时根据信息论是理想调制方法的调制增益（SR_A 是 SR_K的函数）

是关键的因素，它由两部分组成：一直都存在着的量化噪声和由误码而引起的噪声。只有后者与 SR_K 有关。图 1-46 同时给出模拟传输和数字传输的质量与 SR_K 的关系。

图 1-46 有 3 个范围：

a）在接收条件很可靠的情况下数字方法失去优势。

b）一般情况下数字传输比模拟传输更有优势。

c）模拟方法的质量可以无限提高并且超过数字方法。所有的数字方法都有一个质量上限，这是由不可避免的量化噪声引起的。

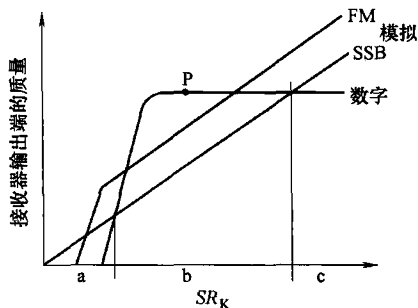


图 1-46 接收器输出端信号的质量与接收器输入端信噪比的关系（程式化描述）

所有能放大带宽的调制方法（也适合频率调制和 PCM）的特点是曲线或多或少会出现弯曲，正如在图 1-46 中所看到的那样。因此当 SR_K 较大时会有调制增益，当 SR_K 较小，即接收条件很差时，相反会出现调制损失。

在 UKW 波段上，使用的是区域 b 内很消耗带宽的频率调制（FM），因为现在调制增益，所以质量会有所改善。相反业余无线电爱好者通常喜欢接收来自很远的地方因而很微弱的信号，大致在区域 a 内变化。很正确的是他们使用 SSB 而不是 FM。

数字传输的最优工作点是 P：输出信号的质量不再与 SR_K 有关。发送功率的继续提高是无用的，因为已经能够正确识别和再生所传输的符号。

点 P 是根据发送功率设置的，接收放大器并不能改善信噪比。发送器必须要了解接收器的运行情况。因为多数情况下，连接是双向的，接收器可以通知发送器它的接

收情况。在 ARQ 系统中（即自动地询问“什么?”）可以用证实消息来通知发送器所希望的接收功率。例如根据 GSM 标准，所有无线电话（手机）的发射功率在 20mW ~ 2W 之间都可以由远端控制。这延长了电池的使用寿命，减小了对相邻无线对话的影响（节约频谱资源），同时也减少了电子尘埃。

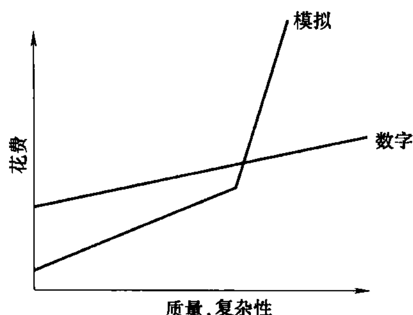


图 1-47　模拟方法和数字方法所要求的质量或复杂性与花费的关系（程式化描述）

图 1-46 中范围 b 和 c 之间的过渡点根据系统的设计而变化。在实践中，可以毫无困难地对数字系统进行设计而使它与 c 无关。例如 CD 无需很复杂就能达到 96dB 的信噪比。使用模拟方法也能达到 120dB，只是太贵，没人会需要它。图1-47针对质量和花费比较了模拟方法和数字方法。数字方法的初始花费较大，但由于符号的可再生性，可以使用合理的费用来实现复杂的系统。相反模拟系统的花费会迅速提高。

结合图 1-46 和图 1-47，可以得到如下的结论：

1）对技术上通常使用的接收信号来说，数字方法要好于模拟方法（图 1-46 的范围 b）。

2）模拟信号只有在以下 3 种情况下还能胜任：

① 需要保证与旧系统的兼容性（例如无线广播）。但目前，尽管费用较高许多地方还是转换到数字无线广播系统：DRM（中波和短波）、DAB（来代替 UKW-FM 无线广播）和 DVB（数字电视）不久就会被引入。具体内容参见 5.1 节。

② 数字方法不受带宽的限制，但因此也阻止了调制增益。

这种情况在电视传输时会遇到。根据表 1-1，TV 信号有 5MHz 的带宽。使用 PCM 这个值会大约增长 14 倍，这样只能容纳少数几个发送站。尽管如此，通过另一种方式也能实现数字方法。带宽与信息量有关，由表 1-4 可以看出，使用数据压缩图像信号的信息量可以减少为 20 ~ 40 倍。之后就可以放心地使用 PCM，因为此时带宽比模拟方法的带宽还要小。当然系统的复杂性会有很大增加，但根据图 1-47 数字方法并不会产生很大的开销。

③ 接收到的信号很弱，但允许接收器的花费不是很大。

这是指图 1-46 的区域 a。上面所介绍的业余无线电爱好者的例子涉及到的就是这种情况。相反，在专业范围内，例如与距离很远的航天探测器的连接，尽管有图 1-46 的结果，使用的还是数字方法。这个图给出的只有调制增益，还有其他方法，如信道编码，相应地将这种方法所得到的质量的提高称为编码增益。但这种方法也会提高系统的复杂性，根据图 1-47 数字方法比模拟方法花费要小。

数字信号的传输和处理不论在技术上还是经济上都特别重要，因此我们仔细要讨论一下。

数字信号在时间和数值上都是离散的。因为信号的处理多数使用的是二进制（两

个值）的数字信号，因此适合光传输。相反在有线传输（电话调制解调器）或无线传输时经常使用多进制数字信号。

二进制数字信号在物理上通常是由脉冲串来表示的，此时脉冲的宽度（持续时间）是常数，脉冲的幅度有两种不同的值。相反，对于理论工作者来说（例如信号处理算法的研究），会将数字信号作为序列，即数字串（例如…0101101…）来描述。

模拟信号能够像 1.1.3 节中所描述的那样，通过采样、量化和编码转变成数字信号。在传输理论中也称为 PCM（脉冲编码调制，参见 3.3 节）信号，这无非就是持续的 A/D 转换。

许多信息从一开始就是数字的，如所有的文本文件、控制命令等。这些信息同样也要用脉冲串来表示。例如将文本用 ASCII（American Standard Code for Information Interchange）码来描述。这正是数字技术的优点，所有信息都采用统一的方式来描述，并且能够使用相同的技术来处理，如图 1-48 所示。

如图 1-48 所示，它涉及从信源符号到传输符号的转变（这也称为映射），此时作为中间产品的始终是二进制数字信号。各种处理使用的都是这些信号，如信源编码（信源适配）、加密和信道编码（信道适配）。

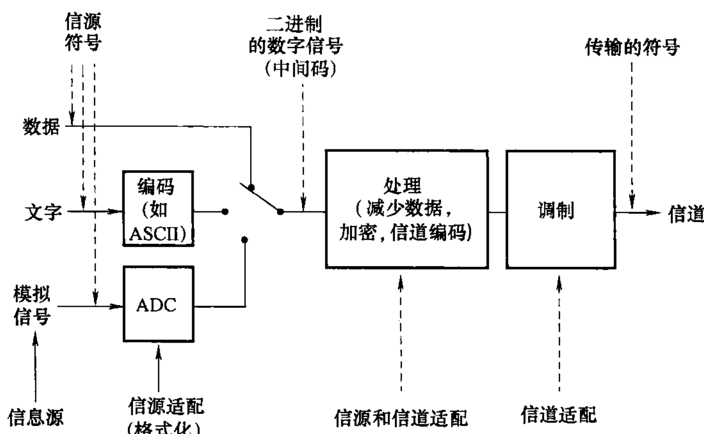


图 1-48 数字传输概述（发送部分，比较图 1-6）

信源符号可为：

1）0-1 序列（数据来自一个二进制信源，如磁盘）。这些符号可以直接拿过来使用。

2）字母、数字和特殊字符（文本、数字源）。这些符号首先必需进行二进制编码，例如使用 ASCII 码。

3）任意的电压值（模拟源，例如传声器）。这些符号必须被分离（采样），并限制在有限的数量上（量化），然后用二进制码来表示（编码）。这正是 1.1.3 节所提到的 A/D 转换。

在数字传输中传输的不同符号的个数是有限的。多种情况都是适合的。例如在基带方法中（详细内容将在第 2 章介绍）

(1) 二进制传输

1) 直方脉冲 5V/直方脉冲 0V（与有/无脉冲的含义是相等的）；

2) 经过平滑的脉冲（平滑会节约带宽）；

3) +5V 脉冲/-5V 脉冲；

4) 20mA 电流/无电流。

(2) 多值传输

1) +5V 脉冲/无脉冲/-5V 脉冲（三进制）；

2) 1V 脉冲/2V 脉冲/3V 脉冲/4V 脉冲（四进制）。

图 1-48 的方案也适合数字信号的带通传输（详细内容将在第 3.4 节介绍），例如：幅度为 A 的正弦波/无正弦波（二进制幅度键控，参见图 1-44 中右），频率 1 的正弦波/频率 2 的正弦波（二进制频移键控），频率 1/频率 2/频率 3 的正弦波（三进制频移键控，参见图 1-9 下），相位与参考相位之差为 0°/90°/180°/270°的正弦波（四进制相移键控）。

数字信号的基带传输和带通传输的主要区别仅在于所使用的传输符号的频谱特性不同。

二进制信号是最可靠的，但为此它需要最多的带宽（抗干扰性和带宽可以互换!）。在带宽不成问题的情况下使用二进制信号，如在设备的内部、采用光传输等情况下。相反在带宽有限的情况下，通常使用多进制信号，例如在无线传输、电话的调制解调器等情况下。

> 在数字传输时，每秒钟要从有限的字母表中传输有限的符号。这些符号的物理特性（幅度、频率范围）必须与信道相适配，并且相互之间的区别要尽可能大。符号的含义要事先与接收器协商。

可以用一个例子来解释清楚多进制传输的概念：如果要传输文本“test”，那么要将它编码成 7Bit 的 ASCII 码的形式：

信源码符号	T	e	s	t
中间符号	0010101	1010011	1100111	0010111

在二进制传输的情况下，只需对符号 0 和 1 给出另外两个符号（有/无脉冲或频率 1/频率 2 等）。

在四进制传输的情况下，需要有 4 种传输符号 0、1、2、3，即 4 个脉冲幅度或 4 个频率：

中间码符号	0010101101001111001110010111
转换码符号	0 2 2 3 1 0 3 3 0 3 2 1 1 3

值得注意的是，一个传输符号可以包含的多个 Bit 中间码（这里为 2 个），这些 Bit 可能来源于 2 个信源符号（字母）。接收器了解传输符号的结构，可以进行方向变换。

在十六进制传输的情况下有 16 个不同的传输符号 0、1、…、15，它们是由 4Bit 的二进制中间码来产生的：

中间码符号 0010101101001111001110010111

转换码符号 2 11 4 15 3 9 7

编码转换仅仅是字母表的变换，见表1-3。

模拟信号传输和数字信号传输之间有着重要的区别：

> 模拟传输时信息隐藏在信号的形状中。因此传输的目标是尽可能不失真地将信号送到接收器中。

> 在数字传输时，符号的形状与信息无关。因此传输关心的只是让接收端了解符号的存在与否。这样无失真是对传输的过高要求（即开销过大）。

> 数字化 = 将形状与内容分离

用户的判决标准也有区别：

> 模拟传输时用户的判决标准是：接收器输出端的带宽、接收器输出端的信噪比（动态）、接收器输出端的线性和非线性失真、花费。

> 数字传输时用户的判决标准是：接收器输出端的传输速率、接收器输出端的误码率、花费。

信道编码能减小误码率（参见4.3节）。为此，了解误码的类型（例如误码是分散开的还是集中在一起的）是有好处的。如果传输1000Bit有3个误码，那么它们出现在第200、500和700Bit还是出现在第499、500、501Bit是不同的。虽然两种情况下的误码率是相同的，但需要信道编码的方式是不同的。

最简单的误码检测系统就是已经提到过的奇偶位：例如一个7Bit的位组需要加长1Bit。第8Bit是这样选择的：总要保证"1"的个数是偶数（even parity）。这样如果有一个Bit发生了错误，即发生了翻转，那么在接收器中校验位的值就与期望的不符，即发现了一个错误，可以使用ARQ系统来重传这个数据位组。但如果在一个数据位组中有2个Bit出现了错误，接收器会误以为正确，没有任何怀疑。

误码的类型对技术人员来说是一个标准，而对用户却不是。在设计巧妙的ARQ系统中，实际上不会有任何误码。在连接状况很差的情况下，接收器很少能正确识别出数据，即接收器输出端的数据速率会下降。

> 在ARQ系统中用户的判决标准是：接收器输出端的传输速率和花费。

1.1.10 电信网概述

到现在为止我们关注的仅是从 A 到 B 单工或双工的点到点传输。在实际中需要的不仅是一个这样的连接，而是同时发生的多个内容相互无关的连接。同时安装多个独立的、相互连接的点到点的连接开销很大，当然是不正确的。多数情况下应该使用网络将几乎相同的用户业务线以很便宜的价格连接起来。这样会充分利用“不是所有人都希望同时通信”这个事实。当然在圣诞节这个假设可能不成立，网络会出现过载，不是每个连接请求都能被满足。

电话网的例子已经控制了许多方面。一部电话机并不意味着已经有一个连接存在，连接必须是拨完对方的电话号码后才能建立。通过建立连接能在网络中预留一条路径，这样通话才可能发生。通信对方是可以选择的，并且一直可以使用同一个终端设备（电话机、传真机等）与这些优点相比，连接建立还有一个小缺点。网络还需要一个能实现交换功能，即临时给一个用户提供一定的传输能力的节点。如图 1-49 所示，从 A 到 B 的逻辑连接被分成多个节点之间的点到点的连接，这些点到点的连接在每个连接建立请求时要重新进行配置。如果没有空闲的传输能力，连接就要被拒绝。因此网络只能提供与使用从 A 到 B 的专用连接非常近似的连接业务。如果不能忍受这个缺点，那就得付出很大的费用来安装一个红色电话（即所谓的专用线）。

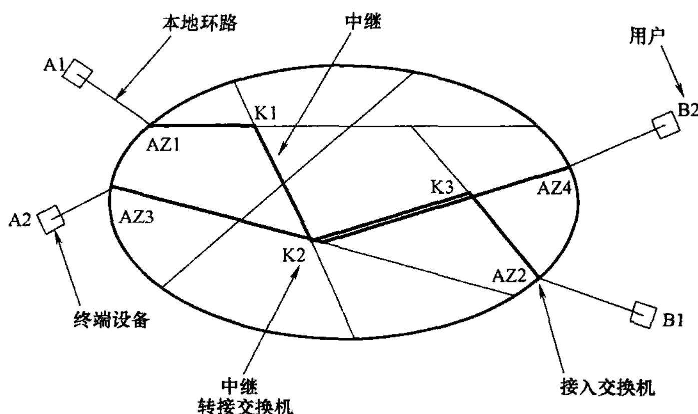


图 1-49 由用户、本地环路、接入交换机、节点（转接交换机）和传输线路（中继）构成的电话网

在电话网中通常将用户描述为用户 A 和用户 B。A 建立连接并且要付费，B 接受连接。连接建立后 A 和 B 具有相同的权利。

每个电话（一般称为终端设备）是通过一条单独的线路与网络相连的。这条线路（并不是终端设备）携带有用户的电话号码，因此被称为本地环路（Local Loop）、用户环路（Subscriber Loop）或用美国的说法为最后一英里（last mile）。所有的接入投资约占电话网总投资的 2/3，因为用户环路的铺设很复杂。本地环路从网络的角度来看终止于接入交换机（AZ），在这里多数情况下信号会被数字化成 64kBit/s。为了节省费用，

AZ 在网络方向上的出口都是经过复用的。图 1-49 中的节点像 AZ 一样都具有交换功能，但它们只与其他的节点相连。

网络内部的传输线路（中继）是功能强大的点到点连接，它们使用光波导(LWL)、定向无线发射、卫星链路和同轴电缆来实现。目前光波导占传输的主要部分。交换机（switch 或节点）将网络的传输能力临时分配给每个用户。复用能降低网络的开销，因为开销不在电子或电缆的材料上，而是在铺设工作上。

因此，电话网由下面的几部分组成：终端设备、本地环路、交换中心、功能强大的复用传输、控制、管理。

网络中有很强的传输能力，每个用户只需要其中很小的一部分传输能力。智能的网络有这样的能力：当一条中继出现问题时，可以为从 A 到 B 重新配置另外一条线路。通过这种办法网络具有所谓的空间多样性，参见 4.3.6 节。

传输技术涉及的是信息的传输。早在 1.1.1 节的开始就已经明确了，信息、能量和物质是相同的基础内容。实际上，观察一下开销较小的传输，这些基础内容都存在类似的问题，解决的办法也是类似的。因此很明显，可以采用相似的办法来工作。

1）在传输信息时会采用光波导或定向无线链路作为传输设备。在传输物质时会使用铁路或汽车，而传输电能量要使用电线。

2）无数的汽车使用相同的街道，此时不同的交通工具被复用在一起。同样也建立了层次化的传输：支路—街道—主干道—高速公路。这些层次化的复用也出现在电力传输中，如低压网、中压网和高压网。

3）在出现瓶颈（拥堵）或部分网络失效（例如施工处）时，可以通过绕行来到达目的地。在传输信息时，可以有大量的绕行线路，网络运营者会留有大量的自由度。相比之下，在传输能量或物质时，基于经济、时间和财政的考虑迂回路径是有限的。

4）当对整个网络的状态了解很深时，网络能够得到充分地利用。因此电信网也为自己的需要传递相关信息。类似地，交通网络的拥堵信息是通过无线广播来传递的。

5）在不稳定的状态下，避免全网瘫痪或能够快速地恢复到正常运行状态下（disaster recovery）是很重要的。尽管会出现瓶颈，业务功能也一定要能够实现。交通网络也是类似的：在出现拥堵时，司机也要给警察和救护车留一条窄路。

上面描述的电话网一百年前就出现了，它被不断地扩建并最终实现了数字化，但在主体构造上它并没有发生任何变化。电话网能够提供世界范围内的人与人之间的话音通信，从 20 世纪 60 年代开始还可以通过调制解调器（Modem）来实现数据通信。

近期为了解决不断增长的数据连接的需求，即从机器（计算机）到机器连接的需求，出现了另一种完全不同类型的网络。这种网络就是现在的局域网（Local Area Network，LAN），它使用一种与电话网完全不同的技术来满足人们对通信的特殊需求（例如客户端—服务器应用）。

不同的相距较远的 LAN 通过远距离网络连接在一起。当然考虑到花费的问题，人们曾试图使用已经存在的电话网来完成这个任务。不过能使用技术上一致的网络来为不同目标、不同数据传输速率、不同安全需求、不同距离等建立连接当然是很好的。目前人们正在为这个目标而努力，但在较长的时间内多种技术还要共存。

上面所描述的电话网络是一个交换网络，而 LAN 是一个广播网络。后者由唯一一条传输电缆（一种类型的总线）构成，所有的用户（计算机、主机）都连接到这条电缆上。此时信息被分成多个被独立传输的分组。接收者再将这些分组组成信息。接收者不断监听信道，并将与它相关的分组从信道中复制出来，而所有与它无关的分组都被忽略。分组中除了有用数据之外还有一个信头，它包含发送者和接收者的地址。因为仅存在唯一的一条信道，所以有可能会出现冲突，单个分组会遭到破坏。使用特殊技术能识别出这种情况并对它进行纠正。在广播网络中用户有这样的感觉：他就是传输介质的唯一用户。

根据地理距离，人们划分出：

1）LAN（Local Area Network），距离为 1～2km（建筑物内，小区域内）；

2）MAN（Metropolitan Area Network），距离为 10～50km 的城市范围；

3）WAN（Wide Area Network），距离可达全世界。

每种网络类型都有它自己的问题，因此都有自己的解决方案。例如大多数 LAN 是广播网络，而 WAN 是通过交换网络来实现的。

从对网络的简短描述中就已经能够看出还有无数的细节问题需要解决。因此在电信技术的发展过程中网络技术已经发展成为一门独立的学科。实际上（很夸张地说）有两类电信技术：

1）传输技术，包括点到点传输的所有技巧，是为整个网络服务的，可以单独对它进行控制。

2）网络技术，点到点传输只是一个功能性的技术，只是出于整体上的考虑才对它感兴趣。总的来说，网络技术考虑的只是发送和接收数据，它只对网络中的行为，即所谓的协议感兴趣。

对两个领域都能很好掌握的全能天才是很少能够遇到的。或许这是因为他们的思考和工作方法是完全不同的：

1）传输技术涉及的是通信的物理方面的内容，因此它需要像傅里叶变换那样的数学概念来描述不同的信号形状。

2）网络技术使用形状完好的矩形脉冲。它只涉及像网络负载、等待时间等这样的基于统计上的理解的概念。通常这个领域不是由电子工程，而是由计算机科学领域的人来研究。

但不合理的是这样的一个事实：这两个领域都很重要，它们甚至相互依赖。同样一个事实是，点到点传输理论已经走得很远了，而在网络方面还在不断出现新的、革命性的思想。电信技术很明显的、巨大的经济意义其实是基于网络。如果只研究传输技术，因特网和它的所有影响都不能够实现。

以前的机械交换机（中继）所使用的是一种与电子方法实现传输（频率复用）完全不同的技术。从 1970 年开始传输就数字化了，对交换机的控制也是一样。但从 1985 年开始信道才数字化。目前从传输到交换的整个路径都实现了数字化，但绝大多数本地环路承载的还是模拟信号。多数情况下 A/D 转换器还放在接入交换机处。引入 ISDN 后，A/D 转换器推广到了终端设备上，现在整个信息传输系统都实现了数字化。

已经能够确定：所有类型的信息都能采用相同的数字方法来描述。这就出现了可用一个统一的网络来承载话音、音频、视频、数据等的思想。ISDN 就是最早的一个代表，即它提供多媒体应用的能力还比较弱。高性能的网络已经可以实现，但其能力还没强到能被广泛使用（如 ATM 参见 6.3.5 节）。

尽管现在信息一直都采用同一种方式来描述，即比特序列，但真正能够综合各种业务的高效网络还不是很简单地就能实现。不同的业务对网络的要求是完全不同的：

1）话音：数据传输速率是常数并且很小。重要的是连续的数据流，偶尔的误码根本不重要。

2）视频：数据传输速率是可变的并且要求很大，单个的误码是可以容忍的，但集中的误码不能容忍。

3）数据：数据传输速率的变换范围更大，但不能出现误码，为此需要有偶尔的停顿。

通过深化一些技术，人们试图能够平衡这些需求，并将开销控制在一定范围内。除此之外还能实现所谓的增值业务（added value），例如网络中的呼叫应答、呼叫转移等。

总的来说已经能够确定：工程师头脑中的关于电信的描述已经发生了转变。以前还严格地区分传输、交换、处理等，现在它们之间的界限已经淡化。网络已经不再由通过中继相连的节点组成，现在网络是一个全分布的节点。数据处理是在一个大型机上来进行还是分布在多个相互连接的小机器上来完成，这种区别已经不再重要了。像 ATM 这种以前实际上是为远距离传输而考虑的技术（传统交换式的网络），现在已经进入到本地网（传统方式的广播网络）中了。

1.1.11 电磁频谱

表 1-10 电磁频谱

波长	频率	英文描述 德文描述		应用
至 1000km	至 0.3kHz	ELF	Extremely Low Freq.	电报 远程打字
100 ~ 1000km	0.3 ~ 3kHz	VF	Voice Frequencies	电话
10 ~ 100km	3 ~ 30kHz	VLF	Very Low Freq. Längstwellen	音乐
1 ~ 10km	30 ~ 30kHz	LF LW	Low Frequencies Langwellen	无线广播、定时广播、导航
100 ~ 1000m	0.3 ~ 3MHz	MF MW	Medium Freq. Mittelwellen	无线广播、航海通信

（续）

波长	频率	英文描述 德文描述		应　用
10 ~ 100m	3 ~ 30MHz	HF KW	High Frequencies Kurzwellen	无线广播、航海通信、飞行通信、业余无线电爱好者
1 ~ 10m	30 ~ 300MHz	VHF UKW	Very High Freq. Ultrakurzwellen	TV、无线广播、航海通信
100 ~ 1000mm	0.3 ~ 3GHz	UHF	Ultra High Freq. Dezimeterwellen	TV、移动通信、卫星、雷达
10 ~ 100mm	3 ~ 30GHz	SHF	Super High Freq. Zentimeterwellen	卫星、定位、导航、雷达
1 ~ 10mm	30 ~ 300GHz	EHF	Extremely High Freq Millimeterwellen	波导、雷达、射电天文
1 ~ 10μm	30 ~ 300THz			光纤传输（红外线）
100 ~ 1000nm	300 ~ 3000THz			可见光

关于符号的解释：HF（高频）和 NF（低频）是德语中超出听力范围以及在听力范围之内频率的集合的概念。英语对应地会使用缩写 RF（Radio Frequencies）以及 AF（Audio Frequencies）。英语的 HF 仅指短波波段。

微波波段（频率超出 1GHz）的带宽太大，因此还要进行细分并使用符号来描述。遗憾的是这些描述并不统一。表 1-11 概括了这些符号描述。可能试图不按照逻辑给它们命名这种做法是很好的。

表 1-11　微波波段的描述（频率单位为 GHz）

L	S	C	X	Ku	K	Ka	E	W
1 ~ 2	2 ~ 4	4 ~ 8	8 ~ 12	12 ~ 18	18 ~ 27	27 ~ 47	60 ~ 90	80 ~ 110

目前这些频率已经变得很紧张了，所以人们试图利用带宽有效的传输方法（当然是数字方法）来减轻这种现象。但这还不够，因此在 ITU（International Telecommunication Union，老的 UNO 下的一个组织，总部在日内瓦）的支持下，每两年要召开一次世界无线电大会（World Radio Conference，WRC）。

在 300GHz 之上的频率还没有正式的划分。

为了能够满足特殊的区域范围内的需求，整个世界被划分成 3 个区域：

1）区域 1：欧洲、俄罗斯、非洲；

2）区域 2：北美和南美洲；

3）区域 3：南亚、日本和大洋洲。

一个很棘手的并且目前很热的话题是电磁辐射对健康的影响。这里对这个话题不应该也不可能展开讨论并得出结论，但也不应该忽略。

所谓的电子尘埃的影响是我们为使用这种技术所付出的代价的一部分，这里具体来说就是提供能量和通信。遗憾的是每个人都想利用它的优点，却不想忍受它的缺点。

按照频率升高的顺序提几个污染的制造者：电力火车、能量传输、家庭和办公室

设备（包括计算机、显示器、电视机）、无线设备（包括无线电话）、发射器、微波炉、雷达设施、X 射线仪器。

这些设备在源处会产生电磁辐射，电磁辐射在暴露的地方会被吸收而产生影响。

尽管目前许多出版物都觉得电子污染无所谓，但忽略这个问题几乎是不可能的。电磁波的影响与不同的因素有关：

（1）频率

1）低频（到几百 Hz）：生理学的影响（影响细胞中的离子交换）；

2）中频：热作用（纺织品变热）；

3）很高的频率（X 射线）：电离影响，对基因的构成产生危险。

（2）功率

在一些实验中功率具有窗口效应，即要产生影响，功率既不能很大也不能很小。另一些实验表明较小的、但是变化的功率比较大的、但保持不变的功率的对人的影响要大。

（3）暴露时间

（4）个人的敏感性（2% 的人群是电子敏感的）

（5）与其他环境和压力因素相结合

持久地暴露在较高的功率密度（例如直接在发射器的周围）下，人们会抱怨头疼、睡眠有问题、免疫系统受到干扰等。

因为多数情况下影响都是很弱的，这个问题很难直接看到并且不能在实验室里进行很好地研究，还不能孤立地分析单个参数的影响。此外人们还增加一些从流行病学的研究所获的经验值，如得癌症的危险。人们不能从一次性地在可追踪到的影响中就得出很危险的结论。另一方面，这个问题在将来肯定不会变小，值得对它进行彻底的研究。

关于吸收电磁波的极限值，相关权威部门已经确定了。但这些值在人们的日常生活中很难根据测量技术来证明。因为没人能给出具体的行为关系，这只是某种程度上歇斯底里的理解。处方不能是不再工作了或在家里不使用任何电能了。通过低频场来测量负荷已经得出结论，在业余时间比在工作时间的负荷小。办公室环境的值要比产品大厅的值小。令人吃惊的是在家里深夜测量到了较大的辐射入侵，原因是枕头附近的收音机闹钟。

注意不要直接在较强的发射器附近。这些设备的操作人员必须要有保护措施。例如观看较高的电缆上的天线馈线，会影响眼中的蛋白质。虽然对大多数人群来说没有必要害怕。但有足够的原因不能将这个问题藏到桌子底下，要进行必要的基础研究。一定要避免的是带有恐惧心理的讨论游戏。

1.1.12 标准和标准化组织

标准在技术上有很重要的意义，它能保证不同厂商的设备之间能够互通。标准也有很重要的经济意义，一定的统一性能保证产品的大规模生产，对减少花费有积极的作用。

因为电信是世界范围内的交易，标准也必须在世界范围内有效。一些特殊的组织

负责建立标准。

标准是合法的、有权威的。除此之外还有建议和事实标准。后者经常是公司自己的、没有权威的标准，但却被广泛使用，因而其他厂商自愿遵守的标准。最明显的一个例子就是 IBM-PC 和它的兼容设备。

标准化组织（负责合法化的标准）可分为国际上代表各个国家的权威组织和自愿组成的组织。代表两类组织的著名实例是：

1）ITU（International Telecommunication Union），UNO 下属的一个组织，总部设在日内瓦（给这个城市带来每4年1次的电信展）。ITU 又划分为

-ITU-T（以前的 CCITT，Comité Consultatif International de Télégraphique at Téléphonique），负责电话和数据交换（以前的 Telex）。在这个组织下产生了例如 V.24（=RS232）和 X.25 标准（见表 1-12）。

-ITU-R（以前的 CCIR，Comité Consultatif International de Radiocommunications），负责无线信息交换。

2）CEPT（Conférence Européen des Administrations des Postes et des Télécommunications）是欧洲国家的电信管理合作组织，它是 ITU-T 下的一个组织。

3）ISO（International Standards Organization）是1946年在自愿的基础上成立的。它的成员是各国家的标准化组织，如 ANSI（American National Standards Institute，是一个私有的组织）。

4）IEEE（Institute of Electrical and Electronic Engineers）是世界上最大的专业协会，拥有重要的 LAN 标准。除此之外，它还制定并给成员提供极好的业务。

5）IEC（International Electrical Commission），主要针对测量设备使用的 IEC 总线等进行标准化。

6）ETSI（European Telecommunication Standards Institute）。

ITU 建议是按系列划分的，以大写字母开头，见表 1-12。到1988年为止，每4年以系列书的方式公开发表新的以及经过修改的建议。系列书以颜色来区分（1980：黄色，1984：红色，1988：蓝色）。因为在这期间标准的范围增长太快，目前建议是单独发行的。

表 1-12 几个 ITU-T 标准的字符含义

字符	含 义	字符	含 义
G	远程电话传输	T	远程信息处理
I	ISDN	V	电话网的数据传输
L	防腐	X	公共数据网的数据传输
P	传输质量	Z	交换的编程语言
Q	信号和交换		

标准出现的时刻对它的成功与否有着重要的意义。当一个概念在实验室中要试用到一个应用中并且进行了细化时，标准化只能是阻碍它的发展。研究阶段之后，到这

个概念在开发室技术上变得更加成熟之前通常有一个喘息的时间，在这个喘息阶段必须要开始标准化。如果开始得太晚，就会为偏离标准太远所做的开发投入过多的金钱，要不断地为实现自己的版本而辛苦的奋斗。如果标准化过早，会阻碍后续的能使技术变得更加成熟的方案。

例如后一种情况出现在无线移动电话系统中。GSM（Global System for Mobile Communication，最初是：Group Spéciale Mobile）是为欧洲的一个共同的目标而开发的。它与以前的模拟系统相比有巨大的优越性。作为 GSM 的竞争者，在时间上稍晚一些，美国和日本开发出了另一种标准，它在带宽有效性上超出 GSM。但用户并没有意识到这一点。多亏时间上的领先，GSM 标准才能被这么广泛地推广，到目前为止它已经被建立起来不能再被封杀。

1.1.13 电信技术的里程碑

1747	实现 3km 长的使用电进行有线传输
1800	发明了电池（Volta）
1835	发明了摩尔斯电码（Samuel F. B. Morse）
1844	在 Washington D. C. 和 Baltimore 之间提供公共电报业务
1858	第一条用于电报的跨大西洋海底电缆（使用了 4 个星期）
1865	提出 Maxell 公式
1866	第一条能够使用的用于电报的跨大西洋海底电缆
1876	发明了电话（Alexander Graham Bell 和 Johann Philipp Reis）
1878	在 New Haven CT 有了第一台交换设备
1887	美国有 150000 电话用户，欧洲有 97000 电话用户
1888	Hertz 证明了存在 Maxell 预言的电磁波
1897	实现了 20km 远无线连接（Guglielmo Marconi）
1899	跨越英吉利海峡的无线连接（Marconi）
1901	横跨大西洋的无线连接（Marconi）
1907	发明了电子管（Lee de Forest）
1920	J. R. Carson 研究了信号采样（没有公开发表）
1933	在 Washington D. C. 实现了第一次高保真传输（Fletcher，Wente 等）
1939	发明了脉冲编码调制（A. H. Reeves）
1947	发明了晶体管（John Bardeen，William Shockley，Walter Brattain）
1948	建立了信息论（C. E. Shannon）
1948	有规律地电视广播（瑞士从 1953 年开始）
1958	第一条用于电话的海底电缆（TAT-1）
1960	发明了激光（Theodore H. Maiman）
1960	第一颗通信卫星（Echo I）
1961	第一条商用时分复用的 PCM 线路（每 24 个信道/1.5MBit/s）
1963	第一颗对地静止的卫星（Syncom）

1965　　商用卫星
1977　　第一个蜂窝移动通信系统（Chicago 的试用系统）
1982　　引入 CD
1988　　AT & T 开始折旧所有的频分复用系统
1991　　GSM 标准开始使用（数字移动电话系统）

电话网的发展在 1950 年之前是很缓慢的，从 1950 年开始不断加速。目前的世界电信与短短 40 年前相比，变化是无法想象的。用户的需求也从电报、电话，发展到无线广播、电视、彩色电视、数据传输，甚至无线电话、全球的交互式多媒体通信。下面通过几个远距离传输的例子来说明电信发展的过程。

跨大西洋电缆

1866 年铺设了第一条从欧洲到美国的跨大西洋电缆，但它只能用于电报。1958 年之前，电话连接是使用短波技术的，质量很差，连接建立过程很笨拙。直到 1958 年，海底电缆对 TAT-1 才开始提供话音信道，但只有 36 个。需要大约 500 个管状、远端馈电的中间放大器，它们之间的距离为 70km。最后一个跨大西洋同轴电缆是 TAT-7，提供 4200 个话音信道，放大器之间的距离是 10km。目前的跨大西洋电缆由光波导组成，放大器之间的距离是 100km，可传输上万个对话。从 1992 年开始也可以用光纤来从一个大洲到另一个大洲传输 TV 信号。但在光纤之前卫星技术已经很成熟，可以在世界范围内提供电视直播。

卫星

1957 年 10 月 4 日，第一颗人造卫星，俄罗斯的斯普特尼克一号（Sputnik）被发送到地球轨道。美国在 1962 年也发射了一颗人造卫星，接着在 1963 年发射了一颗地球同步卫星（Syncom Ⅱ）。1965 年的 Early Bird（晨鸟，又称为 Intelsat Ⅰ，即国际通信卫星Ⅰ，具有 240 个话音信道），成为世界上第一颗实用型商业通信卫星。第一批卫星的生存时间是 3 年，80% 的费用都用于运载火箭。1968 年又向太平洋上空发射了具有 1200 个话音信道和 4 个 TV 信道的 Intelsat Ⅲ，1975 年又有了具有 3780 个话音信道和 2 个 TV 信道的 Intelsat Ⅳ。

1964 年在转播东京夏季奥林匹克运动会时还没有卫星可用，直接发送是不可能的，人们只好乘飞机携带成卷的录影胶片。而仅仅 5 年后，人们就可以惊讶地看到电视中直播的第一次登月的场面了！

光波导

1970 年有了第一条光纤。这种技术上的潜力（低衰减、高带宽、对电场和磁场干扰不敏感、不会产生串扰）很快让人们认识到：15 年的同轴电缆的黄金时间就快结束了。

1986 年 7680 个电话的信号能以 565MBit/s 的速率在一条光纤中传输。放大器之间的距离可达 30km，接合距离是 2km。在相同的传输能力下，1992 年光纤的价格是 1982 年同轴电缆的价格的 1/15。

关于光纤的研究主要集中在相干光传输和光电子上，光传输的速率能到达目前的 10Gbit/s 的多倍，而光电子的功能是将目前用电子电路来实现的功能用纯光学的办法来实现。

1.1.14 电信的经济和社会意义

自从1960年开始，信息论就很大地改变了工业化国家人们的日常生活。这种改变是很迅速的，还看不到发展的终点。电信技术和信息技术已成为很关键的技术。

除了经济意义外，我们还要看一下它的社会意义。每个工程师都要了解它的副作用以及所涉及到的因素（技术的反对者也是如此!）。下面我们就针对这个话题在一定的范围内就个人的观点进行讨论。当然这里并不具有指导性，也不带有哲学和社会学上的意图。它的目的仅是鼓励读者来继续深入地学习电信技术。

人们经常将信息技术或电信说成是第3次工业革命。我们首先将这些概念进行层次化的归纳（见表1-13），然后将这些革命性的影响不带任何观点地进行解释（见表1-14）。最后表1-15给出了何时、哪种技术占统治地位。

表1-13 3次工业革命

序号	时间	发明人	内容	重要的新颖性
1	1764年	Jame Watt	蒸汽机 →工业	通过人为地产生物理上的力，人和动物肌肉的负担被解放了
2	1905年	Henry Ford	流水线 →大规模生产	生产方法比产品本身更重要
3	1947年	John Bardeen William Shockley Walter Brattain	晶体管 →电信技术	信息处理比原材料的处理更加重要

对于3次工业革命时间上的界限，不同的文献资料给出的结果不同。例如作为第3次工业革命的开始有的说是“生产期”（晶体管，1947年），但也有的说是“诞生时刻”（微电子技术，1975年）。

表1-14 3次工业革命所产生的影响

序号	技术创新和影响	社会影响
1	机械化 产品不再受自然(天气)和地理(水路)的影响	农民和手工业者(家庭作业)变成了工厂的工人，工作场所集中化，较长的工作时间和较差的工资(由劳动力的过剩、很大的投资需求和竞争压力而引起)引起社会的危机和不稳定
2	工业化 产品能供给所有的人 变动主要在次要方面(工业)	工人只能看到局部，而不再能看到整体，由此而产生的疏远现在可以通过获得较高的收入来补偿
3	自动化 产品根据服务而改变 刺激第三产业(服务业)	材料的基本需求可以得到满足，相互之间的依赖性增强，环境变得越来越混淆，出现了业余时间组织，提出了生活的意义这样的问题

表 1-15 何时哪种技术占统治地位

1800	1850	1900	1950	2000
蒸汽机 织布机 煤和铁	铁路 电话 摄影 水泥	化学 钢铁 铝 电 内燃机 精密机械	人造材料 电视 航天 电子 核能	微电子 光纤 激光 软件 生物技术
工业化时代				信息化时代

以前，材料运输能力是一个关键的竞争因素，因此国家很努力地修建铁路和公路。目前信息传输的可能性是竞争的一个主要因素，但很遗憾在很长时间内政府对建设电信网的行为需求看到的相对较少。创新的动机和绝大多数投资都来源于工业界。

这种状况改变得很慢。美国 1993 年初开始使用军费来建设做为政府任务的信息高速公路。这种性能强大的、可以传输话音、文字、数据、图像、视频等的数据网（多媒体网络）最先连到学校、图书馆、医院等。从 1994 年开始 EU 也有投资几十亿的大计划。

信息技术的经济利益有多大可以从下面的数字中看出。该数据主要来源于报纸。

1）1995 年对管理人员关于是否对电信满意的调查结果如下：

	欧洲	美国
很满意	7%	37%
不满意	43%	5%

2）在欧洲 33% 的管理人员认为，将来 EDV 家庭办公会增长很快，而在美国有 62% 的人这样认为。

3）1989 年在 BDR（联邦德国）与 DDR（民主德国）之间只有 111 条电话信道，柏林墙倒塌之后的 2 年，就变成了 50000 条。

4）1992 年世界电信市场的总收入大约为 5000 亿欧元。其中的 22% 用于基础建设（投资），78% 用于运行（收费）。增长率是 5% ~ 10%。市场占用如下：美国 35%，欧洲 29%，日本 15%，世界其他国家和地区 21%。

5）信息技术属于世界范围内较大的一个工业分支，规模可以与汽车工业相比。

6）约 50% 的职业会以任一方式使用信息技术。

7）国际电话网是世界上最大的机器，它的替代价值为 15000 亿欧元，其中约 65% 属用户范围（本地环路）。（1998 年）约有 8 亿 8 千万部电话机在运行（美国占 25%，西欧占 31%，日本占 10%，世界其他国家和地区占 34%），其中包括 1 亿 3 千 5 百万移动电话（1992 年只有一千一百二十万移动电话）。仅东京市的电话总数就比整个非洲还要多。

8）欧洲在电话密度上是领先的（1996 年）。瑞典第一，每百人有 68 部电话；瑞士第二，每百人 64 部；德国第三，每百人 56 部；葡萄牙，每百人 38 部。仅美国和日

本两国的电话就可能与除欧洲之外的其他国家和地区的总数差不多。

9）从1975～1995年，国际电话的总数增长了6倍。

10）世界范围内15%的人口占有70%的电信能力。世界范围内通信事业发展最快的中国将改变这种状况。一半的新线路将在亚洲铺设，其中的1/3都在中国。

11）有线电视网已扩展到60个频道，还给中欧安装了150个卫星电视信道。只在1993年就有123颗卫星被送入到地球轨道，其中的50个是通过GUS国家。

根据这些数字，我们就不会感到吃惊：

1）技术是推动创新的一个种可能的因素，但创新最终由市场来决定（由技术推进向市场驱动的转变）。

2）创新是严重的资本密集型的。

电信产业被分成了3个领域：

1）生产商：如Siemens、Alcatel、Ascom等公司就属于这个领域。

2）网络运营商（Carrier）：这些原来属于国家的领域现在发生了很大的变化。除了传统的、现在已经被私有化并被分成邮政和电信两个领域的PTT，目前在这个领域也有一些私有的和国外的公司。图1-50所示为这些公司的规模。由于投资需求巨大，行业趋势为大公司才能生存，所以公司间已经开始有合并的趋势。10个最大的公司占有70%的市场份额。实际上这种现象很奇怪：国家的电信垄断是因为多个提供商有助于为最终用户提供物美价廉的电信业务这种说法而被打破的。现在又因为公司越大，越有助于为最终用户提供物美价廉的电信业务这种说法而进行公司的合并或合资。

3）媒体作为传递的内容。

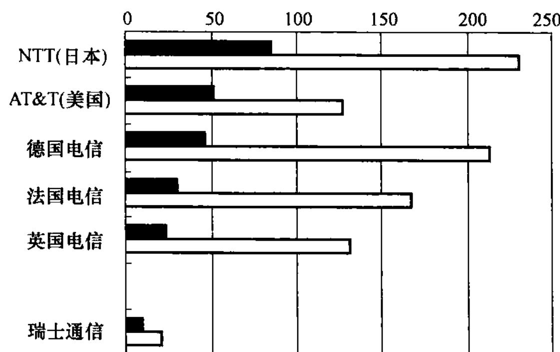


图1-50　1995年以10亿美元为单位的总收入（黑柱线）和以千人为单位的员工数量（白柱线）（上面是5个世界最强的电信企业）

宽带、支持多媒体的网络使用ATM技术是可行的，并且已经能够演示。但它还缺少内容和用户。当然很多用户会吸引很多种类的业务供应，反过来，很多种类的业务供应也能吸引大量的用户，这是一种鸡生蛋和蛋生鸡的状况。希望提供尽可能多的数据传输的网络运营商为了能够拓广传输内容与媒体联系在一起。一个最轰动的例子就是Walt Disney与美国一个大运营商在1995年夏天的合并。例如只有当一个运营商拥有

了电影版权，它才能提供视频点播业务。正因为如此，Microsoft突然对存档感兴趣，目的就是以后能将更多的媒体内容在数据网上提供。对人们日常生活的好的（和坏的）影响都是从内容即媒体而来的。

当然电信也会对社会有影响。为此引用一本杂志上的内容，它抱怨“生活的匆忙、太快的节奏，这就是移动和电信所带来的结果！人们会有这样的感觉，现在在以两倍的速度在生活。与生存的斗争越来越激烈。”另一个引用：“我们变得更快、更匆忙、更紧张。我们现在不用感觉来计算，而用理性和数字来计算。”

经常会听到这种说法。令人吃惊的是，第一个引用竟然出自1906年的瑞士年鉴（作为修饰，用“移动”和“电信”代替“铁路”、“电报”和“电话”），而第二个引用出自1905年的Bürklikalender。如果90年就有相同的抱怨，那么90年的进步带来了什么样的结果？

社会进步的方向是使用文字和图像提供全信息化。但人类本身需要的并不是信息，而是关系。后者是通过信息处理得到的。被淹没在信息中，特别是当它们又缺少关键字时，寻找真正有用的信息是不可解的任务，这反而是反生产的。因此信息提供应该是以信息分发和娱乐为目的的信息灌溉。神学家和社会伦理学家Hans Ruh教授将这种现象称为“Sauglattismus”。除了在相关性和非重要性上的区别，信息的分类或真或假地变得更难了。

技术有时有这样的趋势，将服务者和生产者的角色互换。例如以前汽车使分散化的生活成为可能，但现在分散化却要求有更多的汽车就属于这种情况。同样以前电信使快节奏的生活成为可能，而现在快节奏对电信有更多的要求。这里也不是很清楚，哪些是因，哪些是果。原始的自由和可能变成了必需，这种必需总是到处可见的。是否所有的移动电话真正是必需的还是只是想要的，这是值得争论的。

感谢电信技术使整个世界变成了一个村庄，即地球村，这是加拿大人Mac Luhan所说的话。与城市生活不同，人们在一个村庄里也有很多联系。但也有了较少的个人氛围。因此这只是换了一种颜色或优缺点的权重。地球村使人们的私人氛围减少了，但这出现了另一种程度的文明：在整个世界范围内，人们听同一种音乐、看相同的电视剧因而也收看相同的广告，因此到处都穿牛仔裤，喝可口可乐。

电信使更多的交流称为可能。对另一种文明的匆匆一瞥影响人们的需求，也会使人们对移民做好准备。人们总是在观察，什么事情一直在做，必要时就会采取对策。人们不仅仅只是获得了更多的信息，突然间也可以一起负责。我们能够知悉，还需要做什么，并且已经有过痛苦的经历：作为单个个体我们究竟能够做什么，参见表1-14第三行。

对于关键字“接触”，人类有5种感官可以与外面世界接触：视觉、听觉、触觉、嗅觉和味觉。电信只能满足前两个要求（目前?）。但距离对接收信息（如果狗与狗之间也有电信存在的话，它们也能够传递气味！）来说是最重要的。对整个生活来说，只有两种感官是不够的，因为我们不仅仅要理性地认知世界，还想有意义地生活。电信也有隔离的倾向，这也是人们应该注意的。使用计算机和Modem的家庭工作可能很吸引人，但根据每周与同事亲自接触的时间的长短不同，它也可以是很危险的。同样电

视会议作为在很熟悉的人之间的例行会议已被证明是有用的，但作为第一次相识它却不是很合适。

对于关键字“隔离倾向”，电信不仅能将人联系起来，也能分离人。例如通过为快速的发展节奏所做的努力，它将老人和年轻人分开。它将受过教育和没受过教育的人分开，后者不再有掌控世界的机会。但它也将贫穷和富有分开，因为现代的技术是很昂贵的。这也有个例子：1995 年在日内瓦电信展，这个世界范围内最大的专业展览每 4 年举行一次。1995 年的重点是 ATM（信息高速公路的宽带传输技术）、多媒体和无线通信（卫星和移动电话）。任何人在任何时间、任何地点都是可联系的，但同样在 1995 年，世界人口的 2/3 还从未用过电话。

上面所勾画出的图画根本不是在未来，而是在已经存在的世界上就已经出现了。但只是稍微走过头，还没有完全错误。当然这些发展并不应该只算到信息技术上，这也是更好的交通工具所能提供的可能性、其他科学分支的进步等的结果，简单来说是文明发展的结果。

文明的发展当然不是负面的，而是有些天然的。它不会停止，但希望走很聪明的路。新奇的是发展的节奏以及诸多因素相互之间的关系。因此它首先需要的是一般化而不是特殊化，这样才能在缺口中一次性地打上更多的痕迹。

有这样的说法，为了控制文明的发展方向，“多元化文明”是唯一的处方。为此应该找到一个文明发展的最优值，使用它能使日常生活变得容易，而不是寻找技术上的最大值。毫无疑问信息技术能够并且将对文明的发展起巨大的作用。如（根据 Hans Ruh）：

1）分布化的经济组织，存在兼职工作，并能减少工作时间。

2）开发出新型的并且可靠的安全系统。

3）使用信息传输能限制人员和货物的传输。

4）新型教育系统。

5）提高人的行为能力（研究、医药等）。

因此可以说，不管在何种情况下，对人类本身也有发展的需求，所以值得学习一下伦理学。对工程技术人员来说，参考文献【Stä98】是一本很好的入门书籍。

1.2 信号传输中的失真

1.2.1 无失真传输的定义

信道从电子的角度上可以看成如图 1-18 所示的二端子，而从系统的角度上可以看成是如图 1-33 所示的 LTI。现在我们看一下信号 $x(t)$ 到 $y(t)$ 的传输。这个二端子由放大器、滤波器、线路等来实现。它也可以再安置一个录音机（存储器）。当满足下列条件使，传输被称为无失真或波形可靠：

$$y(t)=K\cdot x(t-\tau) \qquad K,\tau \text{ 是常数};\tau\geqslant 0 \tag{1-35}$$

系数 K 是值为常数的放大系数（在立体声音响设备中 K 可以通过调整音箱的旋钮

来设置，因此这个系统是时变的！），τ 是时间延迟（对于存储器来说 τ 可能是很大的）。因为 τ 一定是正数，因此它是时延，即信号的运行时间。时间向前推移的系统可以（技术上是不可实现的）通过非因果系统来实现。式（1-35）给出的定义相当于无失真的直觉的想像。

在单向传输中（例如无线广播）时延不是很重要的（可以通过传输时间符号来得到）。在双向话音传输（电话）中，100ms 的时延会对对话会产生影响。因此电话系统设置的传输时延极限值是 150ms。对于卫星传输来说甚至允许 450ms 的时延，这当然意味着对会话质量的损害。这是因为会话者与对地相对静止的卫星之间的距离超过 70000km，信号来回所需要的时间已经大于 230ms。因此目前洲际长途电话人们宁愿使用放在海底的光波导。

式（1-35）可以使用傅里叶变换（FT）转换到复数域中。此时 $x(t)$ 与 $X(\mathrm{j}\omega)$、$y(t)$ 与 $Y(\mathrm{j}\omega)$ 具有对应关系。除此之外还利用了线性特性以及 FT 理论中的延迟定理。

$$x(t) \circ\!\!-\!\!\bullet X(\mathrm{j}\omega)$$
$$K\cdot x(t) \circ\!\!-\!\!\bullet K\cdot X(\mathrm{j}\omega)$$
$$y(t) = K\cdot x(t-\tau) \circ\!\!-\!\!\bullet Y(\mathrm{j}\omega) = K\cdot X(\mathrm{j}\omega)\mathrm{e}^{-\mathrm{j}\omega\tau}$$

根据最后一个公式可以计算商 $Y(\mathrm{j}\omega)/X(\mathrm{j}\omega)$，由此给出传输函数以及无失真系统的频率响应：

$$H(\mathrm{j}\omega) = K\mathrm{e}^{-\mathrm{j}\omega\tau} \tag{1-36}$$

$H(\mathrm{j}\omega)$ 是复数，为了方便起见没有使用下划线。换成幅度和相位的形式有：

$$\begin{aligned}&\text{幅度条件：} |H(\mathrm{j}\omega)| = K = const.\\&\text{相位条件：} \arg(H(\mathrm{j}\omega)) = -\omega\tau\end{aligned} \tag{1-37}$$

无失真的因果系统幅度影响是常数，相位响应是线性函数。

一种极端状况是纯电阻网络（例如分压器）：相位响应是常数 0。图 1-51 给出了一般情况。

图 1-51 中的相角 α 满足：

$$\tan(\alpha) = \frac{-\omega_0\tau}{\omega_0} = -\tau \tag{1-38}$$

相位响应的增加相当于通过系统的值为负数的运行时间。

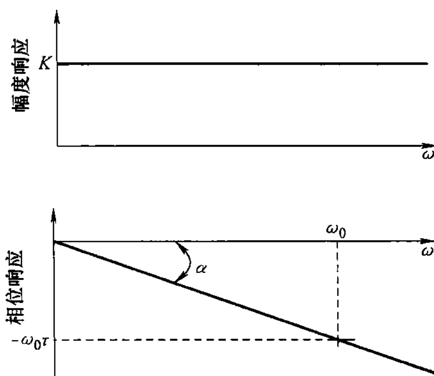


图 1-51　无失真系统的幅度和相位响应

凭直觉就会觉得无失真系统的幅度条件没有任何问题，困难的是相位条件。一个数值上的例子可能是有帮助的。频率为 1kHz 的正弦信号的周期是 1ms，频率为 2kHz 的正弦信号的周期是 0.5ms。如果一个复杂的波形由两个正弦信号的和组成

(傅里叶级数)，这个信号的时延是 0.5ms。因此这意味着 1kHz 信号的周期要减半，即相位要旋转 π。对 2kHz 信号来说要延时一个整周期，即相位旋转 2π。这正好是 1kHz 信号的值的 2 倍，同样频率值也为它的两倍。图 1-52 用图形表示了频率之比为 1:3 的两个正弦信号之间的关系。下面部分两个信号都延迟了 π/2，对一个无失真时延而言，高频信号的时延必需是 3 × π/2。

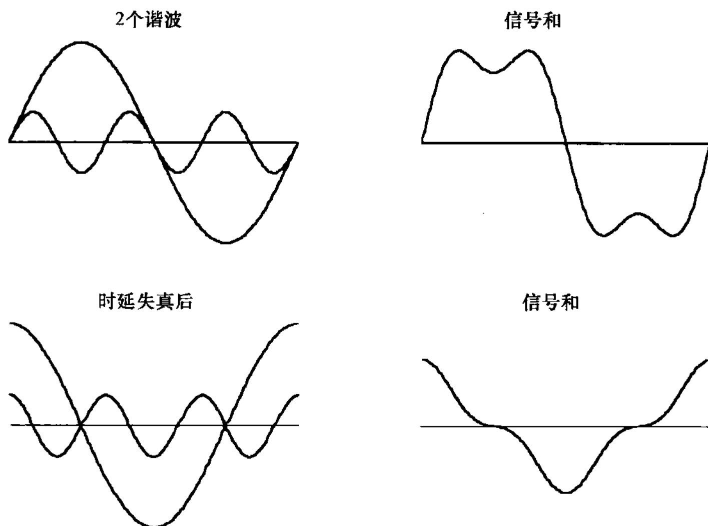


图 1-52 未满足相位条件而产生的失真

注意：一个无失真系统具有线性的相位响应。非线性响应会引起信号变形，尽管系统本身是线性的。对非线性系统而言，根本无法定义频率响应，因此也无法定义相位响应。

对窄带信号可以进行简化：

只要被传输信号的频率成分部分满足条件式（1-37）即可。

例如立体声音响设备中对 20Hz 以下的频率绝对不会实现无失真传输，但这对应用来说根本不重要。

这里还需要强调一下物理学上的波的两个概念：

相位时延：
$$\tau_{Ph}(\omega) = -\frac{\varphi(\omega)}{\omega} \tag{1-39}$$

组时延：
$$\tau_{Gr}(\omega) = -\frac{d\varphi(\omega)}{d\omega} \tag{1-40}$$

在式（1-39）和式（1-40）中相位 $\varphi(\omega)$ 的单位是弧度而不是度。

相位时延在物理上没有含义，由此而导出的相位速度甚至能超过光速。只有在纯正弦形式的系统激励中相位时延才可以精确地给出，并且正好等于信号时延，如图 1-51 所示和式（1-38）。

如果在一个信号中包含多个频率（对所关心的信号来说是这种情况：单频信号是纯谐波信号，即是周期信号，因此不包含任何信息），组时延能够更好地描述系统特性。根据式（1-40）和式（1-37）有

> 无失真的因果系统在所关心的频率范围内，幅度影响是常数，组时延也是常数。

一个常数（即与频率无关）的组时延与没有散射的传输具有相同的含义。

组时延这个概念只有应用到窄带信号（所谓的波组）中才是有意义的。组时延给出这个波组的能量重心的时延。

存在这样的二端子，它在给定的频率范围内具有负的组时延。但这并不意味着这个二端子是非因果系统，它仅仅是指输出端口的包络的最大值比输入端口相应的部分出现得较早。

组时延不应该在物理上一般化地解释成时延。对宽带信号、幅度与频率有关的系统而言，这样做会导致毫无意义的结果。最好是只将组时延作为相位时延的一种测量方法来理解。

1.2.2 线性失真

1. 定义

> 当在传输带宽内幅度条件或相位条件或两者都不能被满足时，系统会出现线性失真。

根据失真的类型分为衰减失真（不能满足幅度条件）和时延失真（不能满足相位条件）。

在因果和最小相位系统（这种系统在右半 s 平面以及 z 平面单位圆之外没有零点）中，幅度响应与相位响应通过 Hilbert 变换联系在一起。因此衰减失真通常也会导致时延失真。非最小相位系统是全通的，系统只影响到相位响应，而不影响幅度响应。

> 线性失真是由线性系统引起的。

> 线性失真的特点是，输出信号中只包含在输入信号中已经存在的频率。这些频率成分通过系统会引起放大、衰减、消失和/或相位旋转。

简单地提前说一下，非线性失真是由非线性系统引起的，在输出信号中会产生新的频率。

线性失真在电线和电缆中会出现，滤波器和放大器也会引起线性失真，反射和多径传输（Multipath）也会引起线性失真，立体声音响设备的音色调节器同样会引起线性失真。绝不是非得要求信号传输一定不能有失真。例如音色调节器表明，失真并不总是不希望的。多数情况下是由信宿（即应用）来决定，失真是可忍受的还是不能忍受的。表1-16给出了人的感觉器官对不同失真类型的敏感性。

表 1-16 人的感觉器官对不同失真类型的敏感性

感觉器官	线性失真		非线性失真
	衰减失真	时延失真	
耳	敏感	不敏感	非常敏感
眼睛	敏感	非常敏感	不敏感

人耳对时延失真的不敏感性当然不是指一个信号的不同成分在完全不同的时间到达人耳这种极端情况。在300～800Hz这个频率范围内，组时延差别升高到20ms都不会对话音的理解产生一点影响。当差别为100ms时，对音节的理解会从80%下降到60%。相位响应对音乐信号比对话音信号重要。

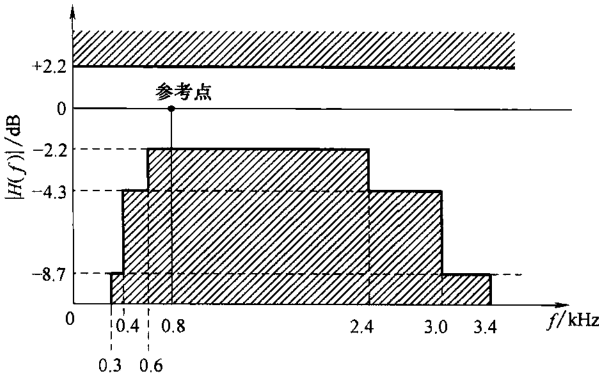


图 1-53 根据 ITU-T 的标准长途电话信道（幅度影响必须在晕滃线内，参考频率是 800Hz）

人耳对非线性相位响应的不敏感性在长途电话技术中已经得到充分的利用。如图 1-53 所示，只需指定电话信道的幅度响应，而无需指定相位响应。相反，图像传输以及数据传输对传输系统的相位响也有着很高的要求，图 1-52 也表明了这点。在这种情况下要使用一个均衡器，参见本节第 3 部分。

2. 用奈奎斯特方法进行测量

一个线性系统可以通过频率响应（幅度响应和相位响应）或是脉冲响应来完全确定。它们之间的关系是通过傅里叶变换来给出的。测量系统可以使用多个方法来实现：

1）使用脉冲序列来测量 $h(t)$，用傅里叶变换来得到 $H(j\omega)$；

2）如图 1-54 所示，使用吞吐量分析仪来测量 $H(j\omega)$；

3）使用噪声激励来测量 $H(j\omega)$（相关性分析）；

4）奈奎斯特方法（段落测量代替环路测量）。

图 1-54 所示为测量频率响应的原理电路。相应的设备被称为吞吐量分析仪，因为在整个相关频率范围内正弦形状的激励信号以很小的步长一点点地增加。分析评估通常采用的是正交相关器。

图 1-54 是经过简化的，因为幅度测量只涉及到输出信号。因此输入信号的幅度必须是常数并且是已知的。复杂的设备也能测量输入信号的幅度，然后求商。相位的测量多数情况下是通过时间测量（两个信号零点之差）然后反推得到的。它涉及的是所谓的环路测量，因为它必须检测输入和输出两个信号。它要求由第二个信道来传输输入信号。这个缺点影响它在电信技术中的使用，因为通常 $x(t)$ 与 $y(t)$ 相距很远。

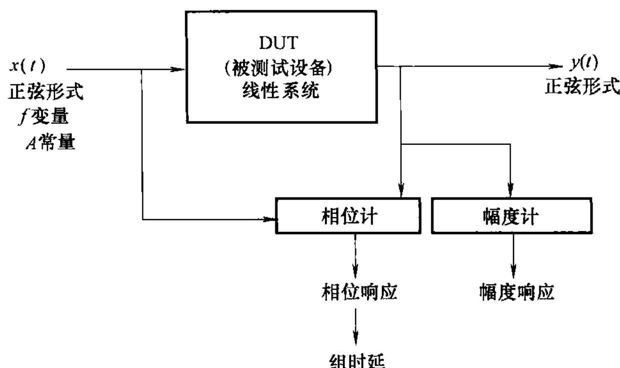


图 1-54 线性系统频率响应的环路测量（DUT）

在低频信道中相位测量可以借助于例如 GPS（Global Positioning System，由卫星支持的导航系统，也可以在世界范围内提供统一的、精度为 100ns 的时间）来实现。借助于所谓的段落测量可以实现修正的奈奎斯特方法。此时只需要检测信道输出端的信号，时间同步可以由被测量的信道来提供。这里需要确定相对于参考频率的组时延，但感兴趣的不是时延，而是时延失真，即时延的变化。放大倍数也不是绝对的，而是相对于参考频率的测量值。

根据式（1-40），组时延是由差分给出的。可以根据差商进行变换：

$$\frac{\mathrm{d}\varphi(\omega)}{\mathrm{d}\omega} \approx \frac{\Delta\varphi(\omega)}{\Delta\omega} = \frac{\varphi(\omega+\Delta\omega)-\varphi(\omega)}{\Delta\omega}$$

上式只有当频率差别 $\Delta\omega$ 很小，并且在这个范围内相位响应是线性时才成立。

使用经过巧妙设计的激励信号就可以产生相互之间离得很近的频率。根据奈奎斯特的建议，图 1-54 中的正弦信号要进行幅度调制，即与第二个带有 DC 偏移的正弦信号相乘：

$$x(t) = \cos(\omega_M t)[1 + m\cos(\omega_S t)] = \cos(\omega_M t) + m\cos(\omega_M t)\cos(\omega_S t)$$

$$= \cos(\omega_M t) + \frac{m}{2}\cos[(\omega_M + \omega_S)t] + \frac{m}{2}\cos[(\omega_M - \omega_S)t]$$

式中 ω_M——测量频率，在所关心的频率范围之内，例如 300～3400Hz；

ω_S——分隔频率，是个常数并且值很小；

m——调制深度，在范围 0～1 内变化。

假设幅度为 1，因为这里主要是对频率合成之后的 $x(t)$ 感兴趣。从上面的计算中可以看出，共出现了 3 个频率，因为分隔频率的值很小，它们之间离得很近，如图 1-55所示。

由于相互之间离得很近，这 3 个频率有相同的衰减和线性的相位推移，即低频谐波与中间频率谐波的相位差等于中间频率谐波与高频谐波的相位差。在时域内，图 1-55 所示的信号有一个正弦信号形式的包络，如

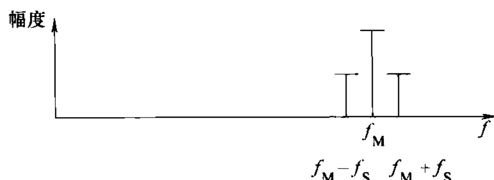


图 1-55 奈奎斯特方法中的激励信号 $x(t)$ 的频谱

图 1-56 上图所示。通过线性的相位旋转，图 1-55 所示波形时域信号的包络同样发生了推移，如图 1-56 下图所示。

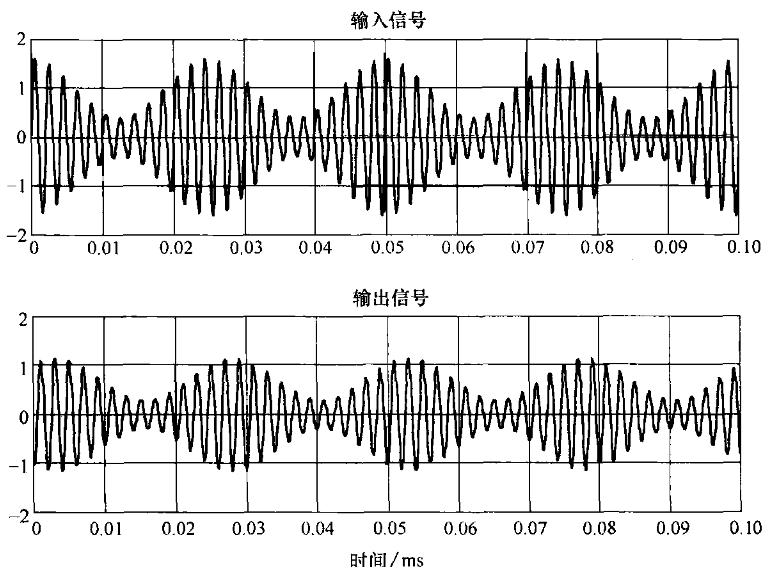


图 1-56 奈奎斯特方法中的输入信号（上）和输出信号（下）

在修正的奈奎斯特方法中，测量频率 f_M 周期性地切换到参考频率 f_R 上。组时延的变化，即所寻找的时延失真，出现在从参考频率切换到测量频率所产生的相位跳变上，如图 1-57 所示。从由参考频率到测量频率这个范围内的幅度还可以确定衰减失真。这种测量方法的一个优点是所有的值都是从接收信号中推导出来的（段落测量），因此这种方法对测量记录设备的线性失真也是适合的。采用修正的奈奎斯特方法工作的设备由工业界来提供。

3. 对策：线性均衡和回波补偿

线性信道引起的线性失真可以通过线性补偿网络来消除。原则上这种均衡器的频率响应是信道的频率响应的倒数，可以简单地放在信道之后。信道和均衡器一起可以看成是一个新的、没有失真的信道，尽管均衡器是接收器的一部分，如图 1-58 所示。

信道可以描述成频率响应为 $H_K(j\omega)$ 或传输函数为 $H_K(s)$ 的 LTI 系统。均衡器的传输函数是它的倒数，即

$$H_E(s) = \frac{1}{H_K(s)}$$

因此级联后的传输函数是 1，可以保证无失真地传输。但这种方法有点太粗糙了：例如 1kW 的发送功率不一定在均衡器中还需再生成 1kW，它涉及的是均衡而不是去掉衰减。均衡器本身允许有一个与频率无关的衰减，同样也允许有一个与频率无关的时

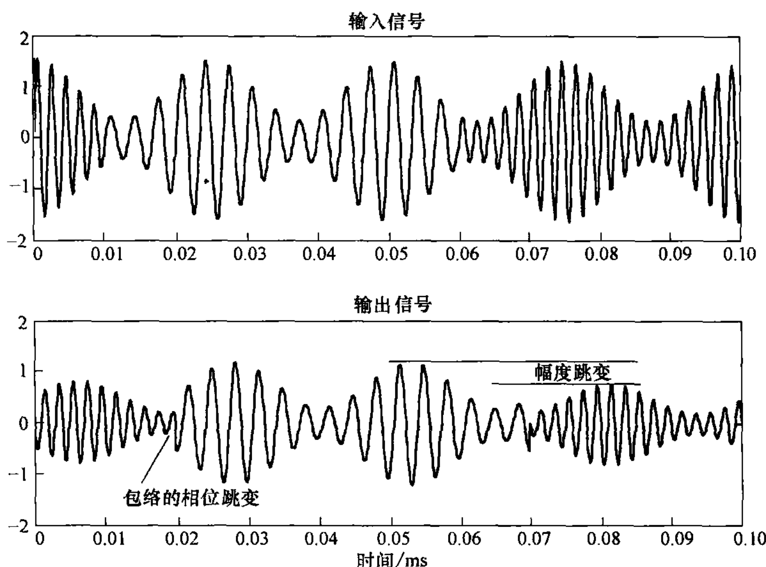


图 1-57　修正的奈奎斯特方法中的输入信号（上）和输出信号（下）

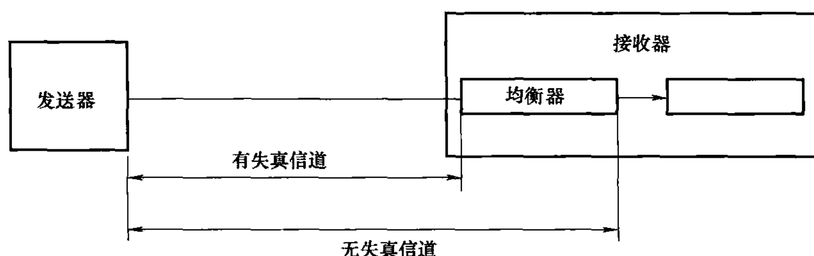


图 1-58　带有均衡器的传输系统

延，因此有

$$H_E(s)=\frac{K}{H_K(s)}\cdot e^{-s\tau}\qquad H_E(z)=\frac{K}{H_K(z)}\cdot z^{-\frac{\tau}{T}}\tag{1-41}$$

式（1-41）左边描述了模拟均衡器的特征，右边是数字均衡器的特征。τ 是延迟时间，T 是均衡器的采样间隔，这里 τ 必须是 T 的整数倍。

但还存在两方面的难点：

1）信道的极点成为均衡器的零点，反过来也如此。如果信道的零点出现在右半 s 平面或 z 平面单位圆之外，这会使均衡器的极点出现在右半 s 平面或 z 平面单位圆之外，这样会引起均衡器的不稳定。但无源信道通常是最小相位系统，即在右半平面没有零点。除此之外，通常情况下部分均衡就足够了。如果信道的极点比零点多，均衡器是不稳定的，并是非因果系统。当 FIR 滤波器作为均衡器时，作为对策可以在 $z=0$ 处增加任意多个附加极点。

2）信道的频率响应不是总是已知的。可能需要在传输真正的信息之前，使用训练序列来测量，然后正确地设置均衡器。有时信道的特性会发生改变，改变甚至会出现在连接建立后（例如利用短波来传输）。为此，必需不断地调节均衡器的值（即可适配均衡器）。

并不是所有的信息传输都需要均衡，见表1-16，但数据和图像传输特别需要均衡。

线性均衡器也能抑制例如由多径传输（无线广播）或混合非对称电路（电话，参见图1-28）所引起的回声（这同样是线性失真）。

均衡器可使用无源电路（例如T形桥）或有源系统来实现。有些电路同时影响幅度响应和相位响应，而全通器只影响相位。通常可使用两级的方法来简单确定均衡器的值。下面将介绍数字均衡器的确定方法，但为此需要对信号处理有一些较深入的了解，可参见参考文献【Mey02】。

图1-59所示为数字均衡器的结构图，作为例子也标出了每个方块的脉冲响应。

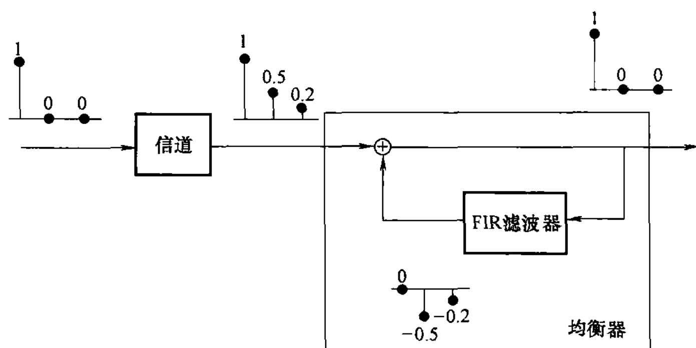


图1-59 数字均衡器

信道的脉冲响应如图1-59中所示，例如为 $h_K[n]=[1\ 0.5\ 0.2\cdots]$，这样通过FIR滤波器，会与序列 $[0\ -0.5\ -0.2\cdots]$ 相加。所产生的和就是理想的无失真的冲激响应 $[1\ 0\ 0\cdots]$。这个和又作为FIR滤波器的输入序列，它的冲激响应正好是所寻找的被加数。从冲激响应的采样值中可以根据式（1-42）直接获得FIR滤波器的系数 b_i。图1-60所示为阶数为 N 的FIR滤波器。

$$b_0=0 \qquad b_i=-\frac{h_K[i]}{h_K[0]};\ i=1,\cdots,N \tag{1-42}$$

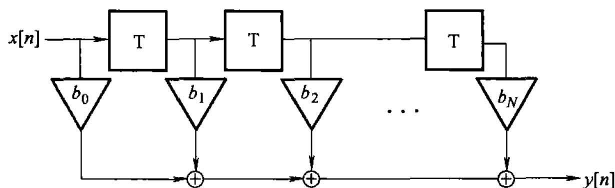


图1-60 FIR滤波器（横向滤波器，Transversal filter）的结构

根据图 1-60，FIR 滤波器的传输函数为

$$H_{\mathrm{FIR}}(z)=\sum_{i=0}^{N}b_i\cdot z^{-i}=\frac{\sum_{i=0}^{N}b_i\cdot z^{N-i}}{z^N}$$

因此反向耦合电路（图 1-59 所框出来的均衡器）除了 z 平面的原点之外只有一个极点。这个电路也可以使用纯递归结构来实现并且有下面的传输函数：

$$H_{\mathrm{E_{rek}}}(z)=\frac{1}{1-\sum_{i=0}^{N}b_i\cdot z^{-i}} \tag{1-43}$$

这里 b_i 又可以通过式（1-42）来获得。作为第 3 种方案，图 1-60 也给出了一种纯非递归的结构。但此时系数需要通过由式（1-43）获得的系统的脉冲响应的采样值来获得。当然这只有当信道的脉冲响应有衰减时才能工作。图 1-61 给出了 3 种均衡器的结构。图 1-62 和 1-63 是图 1-61 中的信号。

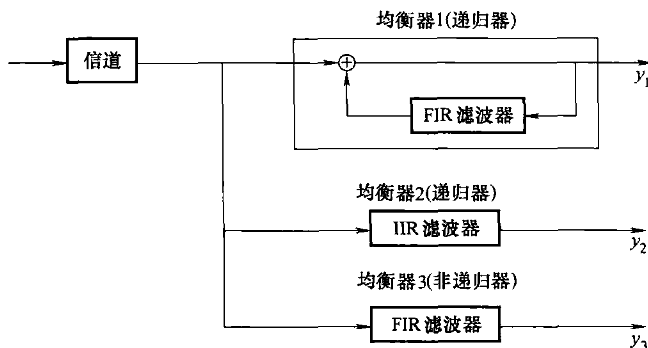


图 1-61　3 种不同的均衡器结构（其中两种递归型均衡器经过简单计算来源相同，因此是等价的）

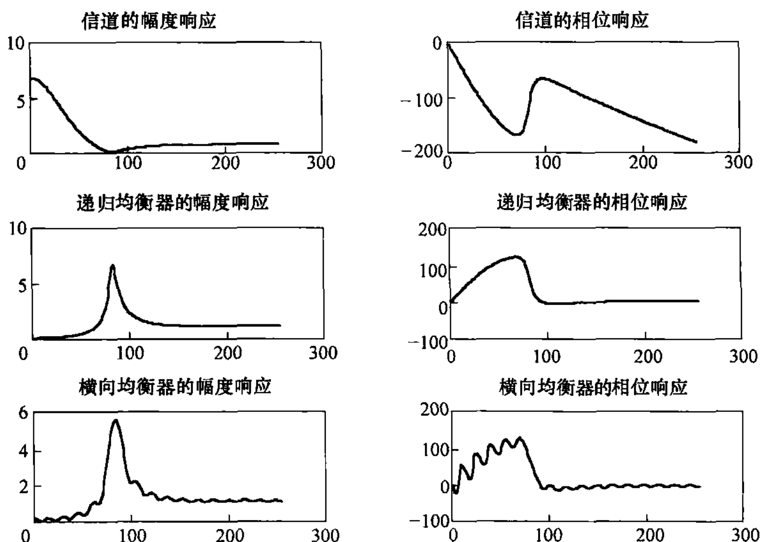


图 1-62　图 1-61 所示信道和均衡器的幅度响应和相位响应

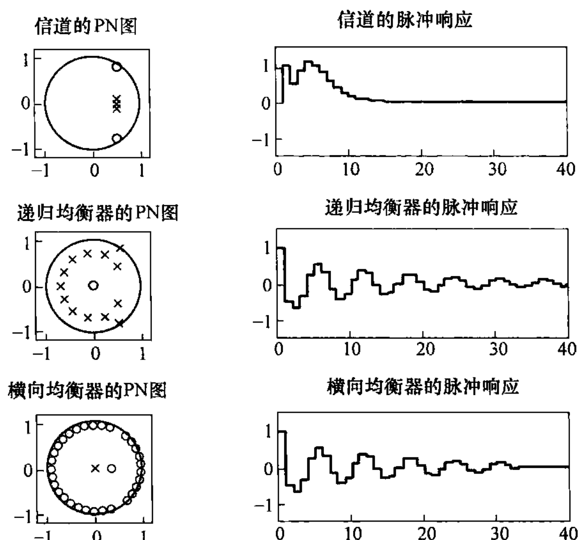


图 1-63 图 1-61 所示信道和均衡器的 PN 图和脉冲响应

图 1-63 右下是横向均衡器的脉冲响应，它的阶数选择得太小。如果将它提高，图 1-62 左下的幅度和相位响应就会变得跟递归均衡器一样好。

图 1-63 左下的 PN 图给出对 FIR 滤波器而言很典型的在原点处的多个极点。只有零点是可以推移的。信道的极点通过均衡器的零点来补偿，但信道的零点是通过均衡器零点环来补偿，这个环在信道零点的位置有一个缺口。

递归均衡器由 FIR 滤波器进行反向耦合而构成，它有可以自由选择的极点，并在原点处有多个不能移动的零点，如图 1-63 左所示。相应地它用极点来补偿信道的零点，用极点环来补偿信道的极点，缺口在信道极点的位置上。

图 1-60 所示的横向均衡器参数的自由度仅与阶数和实权重系数 b_i 有关。可适配均衡器可以自动调节这些参数。优点是 FIR 滤波器永远不能是不稳定的，但确定参数的算法不一定总是收敛的。

横向均衡器也可以在高频范围内实现，时延单元可以通过各线路段或网络采用无源的方式实现。

上面所推导出的均衡器均假设信道的脉冲响应或频率响应是已知的。在实践中这些信息需要通过测量的办法得到，例如在安装设备时或连接建立时。当测量受到噪声的干扰时，必需要对频率响应进行估计，这需要统计信息论的知识（评估和鉴别）。

1.2.3 非线性失真

1. 定义

> 非线性失真由非线性系统产生，会出现新的、在输入信号中不包含的频率。

在实际中实现线性系统时会遇到不希望有的非线性失真，例如对放大器的过分控制、铁心的过饱和等都可以引起个问题。

非线性系统的数学描述是很复杂的，因为叠加原理不再适用，因此不能使用像频率响应等这样的概念。多数情况下需要通过数值计算的方式来在时域内获得系统响应，当然这要通过软件包来支持。

非线性系统的情况下不能通过使用特殊的输入信号（Dirac 脉冲）来得出系统的一般情况，而是要针对每种情况单独进行讨论。结论：

> 非线性系统的测量如果没有精确、确定的测量条件，既是不可理解的也是不重要的！

作为例子我们思考一个二次的特性曲线（例如场效应晶体管）

系统特性曲线：　$y(t)=ax(t)+bx^2(t)$　(1-44)

系数：　a,b 为常数

说明：

1）可以将非线性系统的特性曲线描述成一个幂级数。这种方法在实践中被证明是很有价值的，对特殊情况要使用特殊方法。

2）使用无量纲的信号来计算。在非线性接入时有下面的关系：$[y]=A$，$[x]=V$，$[a]=A/V$，$[b]=A/V^2$。

现在系统使用谐波作为激励：$x(t)=\hat{X}\cos\omega t$

作为响应有

$$y(t)=ax(t)+bx^2(t)=a\hat{X}\cos\omega t+[b\hat{X}^2\cos^2\omega t]$$

$$=a\hat{X}\cos\omega t+\left[\frac{1}{2}b\hat{X}^2+\frac{1}{2}b\hat{X}^2\cos2\omega t\right]=\frac{1}{2}b\hat{X}^2+a\hat{X}\cos\omega t+\frac{1}{2}b\hat{X}^2\cos2\omega t \tag{1-45}$$

最后正好是输出信号的傅里叶变换，尽管这个表达式通过三角函数变换也能得到。新的频率（DC 和 2ω）是可以直接看到的。这个输出信号是周期信号，它的周期与输入信号的周期相同。在稳定系统中这种现象总是可以做到的。但输入端口和输出端口信号的形状是不同这点很重要。

结果的一般结构可以描述如下：

> 如果用只有单个频率 ω（周期 $T=2\pi/\omega$）的谐波作为非线性系统的激励，会产生周期为 $T=2\pi/\omega$ 或是该值的一部分的输出信号，即输出信号的频率为 0、ω、2ω、…、$N\omega$，这里 N 的非线性的深度。

出现的最高频率由 Moivre 定理以及下面的关系来确定：

$$(e^{j\omega t})^N=e^{N\cdot j\omega t}$$

据此可以直接看到幂级数描述的有用性，并且可以有意识地充分利用这些所谓的谐波：倍频（微波技术）和系统识别（为了建模而确定真实系统的阶数）。

根据表1-16人耳对非线性失真是很敏感的。哪些非线性失真的范围是允许的，要由应用来确定。对话音传输而言，那些对音节的理解不产生影响的较大的非线性失真是允许的。但对音频传输而言，很轻微的非线性失真都会造成干扰。

所有非线性失真的测量方法都基于对由非线性而产生的新频率的幅度的确定。通常情况下检测信号的频谱是没必要的。根据不同的应用情况，会采用简化的（因此开销很小）方法。下面将介绍其中的3种方法：

1）失真度测量：对宽带基带传输系统（主要是音频系统）是适合的。

2）双频法：适合于窄带TP或BP系统。

3）噪声失真测量：适合于宽带BP系统。

2. 失真度测量（单频测量）

失真度k是关于由非线性所产生的谐波的强度的一种和的表示方法，因此是表示非线性的一种手段。失真度的定义为

$$k=\frac{\text{从 2 阶谐波开始输出信号的有效值}}{\text{从 1 阶谐波开始输出信号的有效值}}$$

这是在纯正弦波的系统激励的情况下！因此失真度测量也被称为单频测量。

说明：

1）周期信号的基波是1次谐波，谐波频率是基波的整数倍。

2）正确的说法应该是谐波而不是波。谐波只是与时间相关的函数，而波是与时间和地点有关的函数。后者所关注的是传播效果，而前者在高频的情况下很有权威。

3）失真度的定义只涉及到输出信号。因此有可能会存在对系统进行放大而不影响失真度大小的情况。

4）在表达式中没有出现“总信号有效值”的说法，因此可能的DC成分是没有关系的。

5）失真度是对音频技术（这里没有任何DC成分）这种应用进行优化的。

6）参考值是总信号（不包括DC）的有效值，而不单单指基波的有效值。前面这种方法用测量技术较容易实现。

7）失真度用dB或%给出。$k<1\%$（-40dB）能够满足HiFi标准。

8）因为叠加原理不再有效，必需准确地指出测量条件（激励频率和控制方法）。在音频应用中对正常的输出功率来说1kHz的激励频率是很常见的。作为激励信号，需要一个特别纯的正弦波。

9）因为在正弦激励时只能产生输入频率的整数倍的信号，因此失真度只是对宽带基带系统才是很重要的。除此之外激励频率必须很低，否则在允许通过的范围内的谐波会太少，结果会被美化。使用1kHz的激励信号来测量电话信道（带宽为3.1kHz）是没有意义的，但测量立体声音响设备（带宽是20kHz）却是有意义的。

由于三角函数的正交性，可以通过单个频率成分的功率和（Parseval定理）来计算

信号的功率。然后可以给出失真度（U_n 是 n 阶谐波的有效值，由于有断点，也可以用峰值来计算）：

$$k = \frac{\sqrt{U_2^2 + U_3^2 + U_4^2 + \cdots}}{\sqrt{U_1^2 + U_2^2 + U_3^2 + U_4^2 + \cdots}} = \frac{\sqrt{U_{\mathrm{effAC}}^2 - U_1^2}}{U_{\mathrm{effAC}}} \tag{1-46}$$

单个谐波的相位对失真度的大小是没有影响的！图 1-64 是失真度测量的原理图。相应的设备被称为失真度测量桥，它也包括纯正弦波发生器。

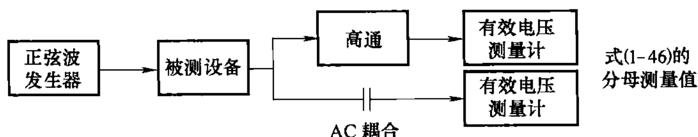


图 1-64 失真度测量原理

3. 中间调制测量（双频测量）

中间调制测量适合于窄带 BB 和 BP 信道。激励信号由两个谐波组成：

$$x(t) = K_1\cos\omega_1 t + K_2\cos\omega_2 t \quad 令\ \omega_1 < \omega_2 并且\ K_1/K_2 = 4 \tag{1-47}$$

系统的特性曲线跟在式（1-44）中一样用指数级数来表示。对输出信号 $y(t)$ 来说，下面的关系成立（N 是非线性系统的深度）：

$$y(t) = a_0 + a_1x(t) + a_2x(t)^2 + a_3x(t)^3 + \cdots a_Nx(t)^N \tag{1-48}$$

如果将 $x(t)$ 用式（1-47）所给出的两个信号来代替，那么传输信号就是两个输入频率的线性组合以及它们的谐波，即在下面频率处会产生谐波

$$f = |\pm i \cdot f_1 \pm j \cdot f_2|;\ i,j = 0,1,2,\cdots;i+j \leqslant N \tag{1-49}$$

将这种效果称为中间调制，将产生的频率成分称为中间调制产物或混合产物。所产生的频率能用所谓的频率金字塔来描述，如图 1-65 所示。

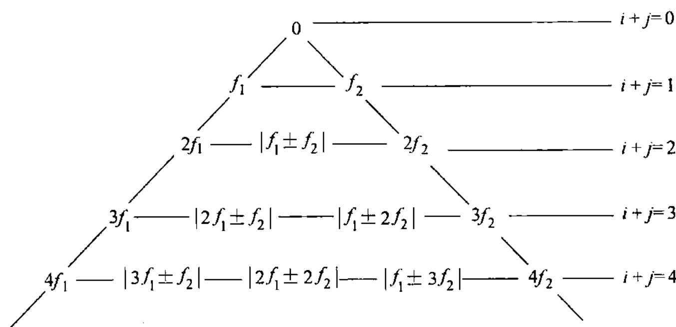


图 1-65 观察中间调制产物的频率金字塔

其中比较重要的是 3 级调制产物（即 $|i| + |j| = 3$），也就是频率 $2f_1 - f_2$ 和 $2f_2 - f_1$，原因如下：

1）非线性系统特性曲线通常是准对称的（反时钟电路），即式（1-48）中的偶数

系数很小，相应地偶数阶中间调制产物的值也很小。

2）随着阶数的升高幅度是下降的。

3）占统治地位的新频率因此是奇数阶和较低的阶，也就是说是第3阶。此外这些成分还在可通过范围，不能被过滤掉。例如：可通过频率范围是100~110kHz。如果f_1 = 104kHz，f_2 = 106kHz，那么$2f_1-f_2$ = 102kHz。

除此之外，中间调制是建立短波接收器时会遇到的主要问题：通常会遇到这种情况，在较小的频率间隔内的很弱的信号应该被特别强的发送器（例如无线广播）所吸收。弱信号对自己来说还是足够强的，足够能被接收器检测到。但它还是因为较强的3阶中间调制产物而被两个邻居无线广播发射站所覆盖。这个中间调制在接收器中由有源部分（输入放大器和混频器）来产生。解决办法很简单，但第一眼看起来很荒谬：采用无源（并且因此是线性的）衰减器将所有信号都减小例如20dB，这样3阶中间调制产物会减少甚至60dB，这样就不会覆盖希望接收到的较弱的信号了。

4. 噪声失真测量（多频测量）

这种方法适合于宽带BB、TP和BP信道，但在实际中仅适合于BP信道，因为对BB和TP信道而言，失真度测量方法更简单、更便宜。宽带BP信道会涉及到例如频分复用系统。由于有放大器等，这种系统会有些非线性的特点，会在同一频段内产生新的频率，这些频率会与复用频段内的其他信道的话音产生串扰。

图1-66所示为噪声失真测量的主要原理。宽带噪声信号（它们与例如500个经过复用的话音信号具有相似的特性）作为测试设备的激励，从带阻滤波器中所关注的频段内截取几个间隔。在非线性系统中由于中间调制会产生“无限”多新频率，其中的几个会落到这些间隔中。在系统的输出端使用带通滤波器，因此只有之前的几个间隔会被过滤出来，可以测量其中所包含的噪声功率。这个功率越小，被测设备的线性就越好。前面提到了单频和双频测量方法，噪声失真测量也被称为“多频法”。

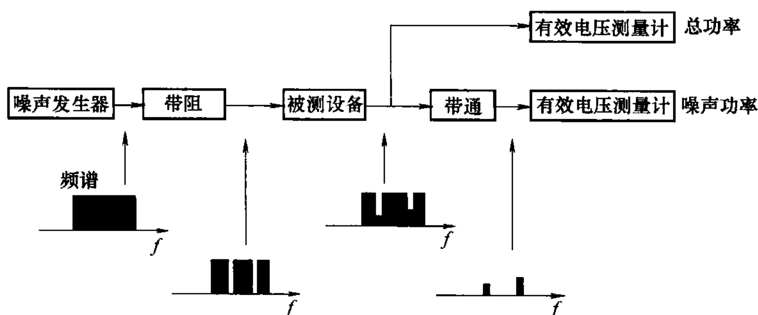


图1-66 噪声失真测量的原理

5. 对策：非线性均衡

对非线性系统的均衡使用的同样是非线性均衡器，系统实现的是均衡器的倒函数。这只能是当均衡器具有明确的结构时，即倒函数是可描述的情况下才是可行的。例如比较器不能被均衡。相反如果失真系统具有对数系统函数，那么可以使用指数函数来

进行补偿。我们在 1.3.2 节开始部分中又会遇到这种情况。

1.3 信噪比的改善

从式（1-17）可以看出，对一个成功的信息传输系统而言，除了带宽之外，信噪比以及动态性也是很重要的，因此需要会使用特殊的方法来提高信噪比。

1.3.1 线性方法

1. 过滤

信噪比是有用信号功率与噪声功率的比值。这个比值可以如下的方法来提高：提高有用信号功率（提高发送功率，充分利用天线的方向性等）和减小噪声功率（屏蔽噪声源）。

过滤遵从的是第二种方法，即通常情况下遇到的噪声信号的带宽比所关心的有用信号的带宽要大。在接收器输入端使用线性滤波的方法会减小噪声功率而不影响有用信号的功率，这样信噪比会得到提高，如图 1-67 所示。在实际中每个接收器的输入端都有一个这样的滤波器。

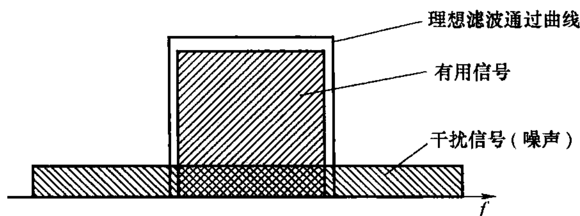


图 1-67 有用信号和噪声信号的频谱

如果带宽为 B_N 的噪声频谱的曲线是常数，那么过滤后信噪比会提高 B_N/B_S 倍（B_S为有用信号的带宽）。

2. 预修正

“提高发送功率”这种方法不总是很实际的，例如因为馈电问题。至少能改善主观（人感觉到的）信噪比的一种方法被称为预修正。在图 1-67 中，信号和干扰频谱都是常数。但实际并不总是这样的，例如噪声功率可能在高频处会提高，如图 1-68 所示。采用线性的方法来描述有用信号可能会过多地强调它的高频部分而使低频部分付出代价，这样信噪比在发送功率不变的情况下与频率无关。在接收器中使用去修正的方式执行相反的过程。但客观上（通过测量设备来测量）使用这种方法信噪比并没有得到改善。

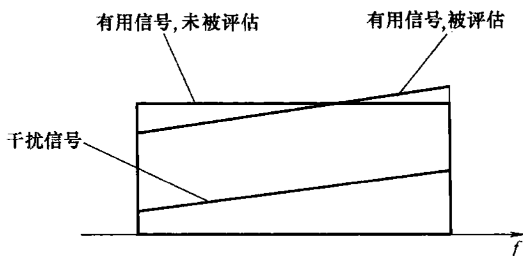


图 1-68 通过预修正改善信噪比

噪声功率的不均匀分布会在例如基于频率调制的传输方法中遇到，也会出现在 UKW 无线广播中，那里当然会使用预修正方法。

信号功率的不均匀分布会出现在使用同轴电缆的宽带信号传输中（电话和有线电

视的频分复用)。原因是集肤效应（电流会集中在电缆的外表面)，频率增大时电缆的电阻会增大。

3. 补偿

最后还可以通过补偿的办法来改善信噪比，前提是能提取出噪声信号。这样噪声信号要么是已知的，要么能单独产生它（例如网络影响、干扰音等）或是能独立测量到它。现在要将干扰信号从总信号中抽取出来，这样在理想情况下就只剩下有用信号了。通常为了补偿会采用相关的方法，即统计信号分析法。最著名的一个例子就是胎儿诊断：未出生孩子的心脏音是很难控制的，因为它会部分地被母亲的心脏跳动所覆盖。使用第二个传声器来单独测量母亲的心脏音，并将这个信号放大并延迟，这样通过取差就能测量到胎儿的心脏音。

在数据传输中使用补偿的办法来消除双工传输（见图 1-29）中由混合电路所造成的回声。图 2-20 所示为回声抵消器的一种实现方法。

1.3.2 非线性方法：音节压扩和瞬时值压扩

瞬时值压扩

噪声信号作为一种干扰，在较短的平均时间内功率是相对恒定的。相反话音信号在一个时间片内幅度可能很大或很小。对后一种情况来说信噪比会很差。一种改善方法是在发送器中使用特性曲线为非线性的（多数情况下是对数的）放大器，相应地在接收器中使用指数放大器（压扩器）来对压缩进行反向操作。

这个瞬时值压扩会引起非线性失真，传输带宽会明显地加大。为了使压扩器能够无差错地恢复出原始信号，在传输过程中不允许有其他失真（特别是延迟失真)。

在电话网中的话音信号数字化过程中会广泛地应用瞬时值压扩，参见 3.3.1 节第 3 部分。采用这种方法量化信噪比会显著提高。

瞬时值压扩只是通用原理的一种特殊情况，通用原理在高频传输中也被使用，即扩频技术，参见 5.3 节。

音节压扩

瞬时值压扩的缺点是会增加带宽，针对这点使用音节压扩会更好。通过调节发送器中的放大系数，可以将输出信号包络的变化减小，这样较小的一段会被提高，因此人们也将这种方法称为动态压缩。只有在瞬时放大系数的信息也被同时传递的前提下，接收器才能进行正确的扩展。这可以通过例如使用唯一所关心的频率范围之外的导频来实现。在传输话音时（例如通过短波）接收器通常不使用扩展，这丝毫不会影响话音的可理解性。以前磁带录音机经常会遇到动态性的问题，因此那里会使用一个音节压扩器（ALC)。但在录制音乐时动态性损失代表质量的损失。

准确地说音节压扩不是非线性系统，而是一个线性时变系统。因为放大的变化是非常缓慢的，在实践中传输带宽并没有增加，因此这种方法也可以毫无问题用在传输系统之后的额外的附件上。

数学上，可以采用下面的方法表示压扩器（x 为压扩器的输入信号，y 为压扩器的输出信号)：

线性系统：　$y(t) = Kx(t)$

时变系统（音节压扩）：　$y(t) = k(t)x(t)$

非线性系统（瞬时值压扩）：　$y(t) = k(x(t))x(t)$

输出信号的频谱可以通过上面系数频谱的卷积来得到。线性系统没有引起带宽的增加，时变特性只增加了很少的带宽（$k(t)$ 变化很缓慢，并且是窄带的），相反非线性系统会大量增加带宽。

第2章　数字信号基带传输

2.1　引言

在第1章中我们已经了解到，信息都是使用数字来描述的，例如一串脉冲，如图1-48所示。本章讨论的主要内容是如何对这样的数字脉冲进行远距离的传输，如图1-22所示。

本章对脉冲串的来源（直接来自信源，或通过信源编码，或来自复用器）和内容（音乐、数字等）并不关注。本章所关注的仅仅是如何使用尽可能小的开销，将尽可能多的脉冲尽可能安全地进行传输。同样是否使用信源或信道编码，是否进行加密，或是否进行复用等在本章中也无关紧要。此外，本章中对信号本身也不进行重要或不重要，冗余或非冗余的区分。

需要提前说明一点：在涉及到通信网相关方面内容时，信息传输的不同行为会根据OSI模型来进行划分，而OSI模型将在6.1节进行介绍。如果需要的话大家现在就可以直接翻到那里进行查找，本章的内容仅涉及OSI模型的第一层。

本章不涉及的内容还包括幅度离散但时间连续（即模拟信号）的脉冲序列，如PWM、PFM等涉及到的情况（如3.2节所述内容）。因对于远距离传输而言，这些信号具有很特殊的含义，具体细节详见参考文献【Mäu91】及【Höl86】。同样，时间离散但幅度连续（也是一种模拟信号）的脉冲，如PAM涉及到的情形（参见3.2节），也属于没有明显涉及到的内容。但本章下面所述理论对PAM信号却是适用的。

我们这里所涉及到的发送器能够发送一串如图1-10下半部分所描述那样的脉冲。接收器则接收到如图1-10上半部分所描述的信号，这些信号经过再生后转发给使用者。信号的再生是通过接收器在每个脉冲的中间点测试输入信号以判断脉冲的存在与否来进行的。因此接收器必须要能够识别时钟的频率和周期，这样它才能在适当的时间点进行测试和判决，也就是说，接收器必需要有一个同步系统。通常人们对同步系统进行如下的划分：

1）位同步：即脉冲的中间点在哪里？

2）字同步：一个数据字从哪里开始，即哪里是MSB？

3）帧同步：即第一个数据字从哪里开始？这对数据网来说很重要，因为数据网会将一长段信息分割成数据包，并且将每个数据包单独进行连接（即分组交换）。其他网络将连续的数据流进行转换，这些数据流可能从多个信道，经过时间复用而得到。在这种情况下接收端必需能够正确识别多条信道。

也可以近似地认为同步由相伴的另外一个单独的连接来实现。如果使用并行传输的话，位同步甚至可以省略掉：传输8Bit的数据字需要10条线，即每个Bit各一条，

一条用于进行同步，而另外一条作为公共回路（参考或测量电压）。很明显这种方法使用的铜线极易腐蚀，因此它仅适用于短距离传输，如从一台计算机通过并行接口连到一台打印机。

对于长距离传输来说，通常使用的是串行传输，所有的脉冲通过同一条线路一个接一个地传。同样也省掉一个单独的同步信道。接收端需要独立地从接收到的数据流中进行位同步和字同步。这个问题主要根据两个原理来解决：异步传输和同步传输（参见 2.2 节）。表 2-1 总结了传输距离与串行、快速传输与同步之间的基本关系。

表 2-1　数据传输方式分类

数据传输	异步	同步
并行	短距离	快速传输，短路离
串行	长距离	快速传输，长距离

脉冲串的频谱服从 sin（x）/x 分布，能量集中部分的宽度随脉冲的宽度（持续时间）而变化。理论上来说所需要的带宽是无限大的，但实际上带宽受信道的限制。这种带宽的限制会影响所传输的信号，最严重时接收端可能不能正确地识别出脉冲。因为信息仅仅是与脉冲的存在与否相关，而不是脉冲的形状，所以在发送前通常对脉冲进行适当的变换，以便它们在传输时需要尽可能小的带宽，并且虽然受信道带宽的限制，但是变形并不是很严重（参见 2.5 节和 2.6 节）。

如果单位时间内要传输的信息内容越多，脉冲的宽度越窄，速率也就越高。但是脉冲宽度越窄，传输所需带宽就越大。这也符合已知的事实，即传输速率越高，所需信道的带宽越大。但频率的下限却与传输速率无关（低通信号）。通常传输信道由于放大或是分压不采用 DC 耦合（比较图 1-22 所示的 AC 耦合的基带信道）。这个问题会通过使用信道编码（参见 2.4 节）来解决。信道编码还有一个任务就是在出现较长的全 0 或全 1 序列时还能保证接收端能恢复时钟，即确保位同步的实现。

信道的影响（时延、回音）会使信号发生扭曲以至于不能被正确识别，均衡和回音抵消负责减少这种影响（参见 2.6 节和 2.7 节）。

信道所受的干扰会使脉冲的形状发生变化，偶尔还会使接收端的探测和判决发生错误。在数字信号的情况下信道噪声不会使信噪比变得更差，但会增加误码率。可以根据概率计算和统计的方法来计算误码率（参见 2.8 节）。

这里还要再回顾一下图 1-24：第 2 章中关于数字信号基带传输的许多原理对 3.4 节所述的数字信号的带通传输同样适用。

2.2　异步和同步传输

接收端必须在脉冲的中间进行判决。虽然正常情况下脉冲的频率可以通过约定来确定，但在实践中，相互独立的振荡器制造并不能做得很精确，经过较长的时间没有几个 Bit 能正常接收（Bitslip）。这种情况跟两个相互咬合在一起的齿轮一样，如果每个齿轮都由相互独立的发动机来驱动，那么经过很短的时间齿轮就会磨损得很严重！

在异步传输时，接收端时钟的恢复虽然与约定的时钟频率无关，但是每个字节过后都要重新同步，这样时钟的周期才能始终保持正确。为此需要将数据流分成多个组，每个组由多个比特组成。通常情况下正好是一个字的长度（如使用博多码（Baudot-code）的电报系统中字的长度为5Bit，而带有奇偶校验的ASCII码中字的长度为8Bit）。每个组外加一个开始比特（StartBit）和结束比特（StopBit）构成一个帧。这两个帧比特的物理状态不同，其中StopBit的长度是任意的，但最小长度是每个比特持续时间的1.5倍，如图2-1所示。

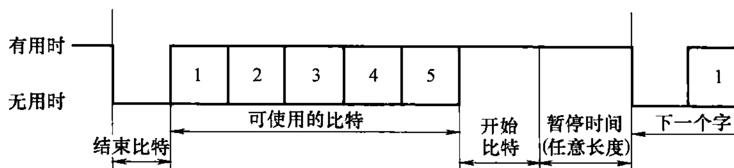


图2-1 异步数据传输（50Band带有电流环路的电报传输）

所有比特，包括帧比特的逻辑状态都要么是0要么是1。物理状态可以是例如0V和5V（TTL信道）、-12V和+12V（V24接口及RS232接口）或是0mA和20mA（电流环路）。在3.4节（数字信号的带通传输）中也使用异步通信：在FSK（频移键控）过程中两个不同的状态对应于两个不同的频率。

接收器来识别时钟的速率（图2-1中每个脉冲20ms宽，时钟频率是50Hz）和字的宽度（在例子中是5Bit）。在任意宽度的停止比特中间接收器搜索下一个开始比特的边缘（图2-1中暂停部分）。开始比特的边缘一旦被识别出，10ms之后立即开始对脉冲进行判决（即在开始比特的中间）。经过20ms再进行一次脉冲的判决，这样的过程还要再重复4次。然后为了进行同步控制还要再对停止比特进行检测。为了能够尽可能准确地确定开始比特的中间位置，要使用更高的频率（例如10倍于）来进行探测，寻找开始比特的边缘，然后再将探测频率恢复到正常值，如图2-2所示。如果接收端的频率振荡器与正常值有所偏差，测试时间点与边缘只有一个脉冲的偏移，也能较好地确定停止比特。这种方法对字长度是任意值的情况也不会产生任何问题。

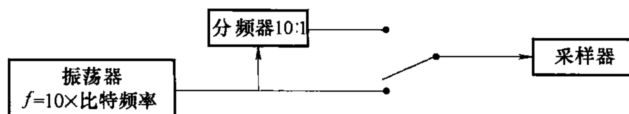


图2-2 异步传输中的同步（在字的检测过程中开关置于上面的位置，在停止比特和暂停过程中开关处于下面的位置）

同步过程一旦建立，接收端的其他过程就很明显了。经过一段时间的中断之后，必须在脉冲串中重新找到开始比特。此时将出现与开始比特的状态（在我们的例子中是没有流）相同的第一个比特作为开始比特，然后根据上面所描述的办法来对整个字进行检测。最后对可能是结束比特的比特的状态进行控制。如果误将字中的比特当成开始比特，那么结束比特的状态就是错误的，这样上面所描述的过程就得重新开始。

此时可能还是刚开始以为找到了开始比特，发现了同步，但到最后还是被结束比特的错误状态而否定了。但总有一次会找到正确的开始比特。其实找到正确开始比特所用的时间并不会很长。由于规定停止比特的时间间隔最短是 1.5 倍的时钟长度，这样对不太有利的有用数据串来说也能很快并且很安全地找到同步。

大多数情况下发送端以并行的方式从数据源那里得到数据。为了实现并行到串行的转换，发送端需要一个移位寄存器、一个时钟发射器和一个控制逻辑。接收端需要同样的设备单元。这样的收发电路可以很便宜地、以集成电路的方式购买到，集成电路的名字叫做 UART（Universal Asynchronous Receiver Transmitter，通用异步收发器）。

异步传输的优点在于同步较容易进行，并且比较容易识别出状态“0”、“1”和“无信号”（因此开始比特并不是比特流）。另外该方法由人工将数据分成所谓的异步数据输入源。正是因为这个原因异步传输几十年来一直被电报所使用。但是一个物理上与电话网完全分离的并且进行 DC 耦合的网络是需要的。目前已经由传真和调制解调器代替了电报。

异步传输的缺点在于数据的吞吐量很小：在上述例子所述的电报系统中，每 5 个有用数据比特需要 2.5 个帧比特。后者同样需要传送，因此该例中的数据吞吐量最高能达到信道传输能力的 66%。如果字的长度是 8Bit，这种关系也不会有多大的改善。因此异步传输是传输少量数据的最简单的方法。

相应地同步传输被快速传输所采用。此时时钟由接收到的数据信号来产生，而时钟再生器是接收器一个必要的组成部分，绝大多数情况下由 PLL（Phase Locked Loop，锁相环）来实现，如图 2-3 所示。

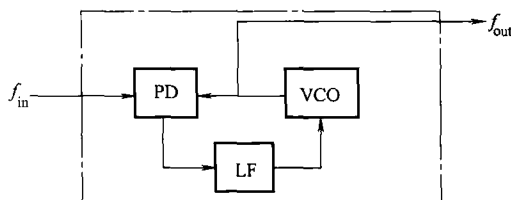


图 2-3　PLL 的组成

PD—相位检测器　LF—环路滤波器（低通）

VCO—电压控制振荡器

在图 2-3 中，如果输入信号与 VCO 信号的相位不一致，相位检测器就会产生一个信号，根据这个信号 VCO 会进行调整。在锁定的状态下，VCO 信号的相位与输入信号的相位有一个固定的对应关系，因此这两个信号就有相同的频率。

多亏有了信道编码技术（参见 2.4 节），使得在接收信号中能够包含时钟信息，如在数据串频谱上或多或少离散地加入些频谱线，PLL 就能够抓到这些频谱。这样它的有规律的动态特性就会被带宽很窄的环路滤波器有意识地限制住。VCO 的输出信号因此具有一定的“飞轮效应”，受短暂的扰动仍能保持产生时钟。

输出信号可以显示在每次环路控制时有多少偏差，也就是说可变的相位偏移。这个所谓的抖动会导致检测点的不稳定性。检测时间点的最大波动被称作抖动幅度，波动的快慢称作抖动频率。

异步传输中的位同步和字同步问题一步就能解决，而在上面所述的同步传输中，通过使用锁相环（PLL）才能保证比特（位）的同步。

字同步需要采用同步字或帧字才能解决。同步字或帧字是一个事先约定好的位图，

如00110011。接收端必须要了解后续的帧的构成，如可以是每511个字后面可以跟一个8Bit的帧字。为了进行字同步接收端必需寻找这个帧字，即数过511×8Bit之后找到想要得到的帧字。

经过很短暂的连接中断之后接收器开始寻找帧字。由于事先规定好的位图在有用数据当中也会出现，接收器可能会误将这些数据当作帧字来处理。经过一帧之后接收器会意识到这个错误，除非此时正好在有用数据当中又出现规定好的位图。只要在有用数据当中不周期性地出现这个位图，早晚会实现帧同步。

为了保险起见，可以使用比特填充的方法，以避免在有用数据当中出现事先规定好的帧字的位图。例如如果选择01111110作为帧字，那么有用数据就得这样修改：经过连续5个1之后，就得填充一个0。在接收端，如果经过连续5个1之后出现0，就得将这个0去掉。这样在接收端只要连续出现6个1，就可以按照帧同步字来处理。

在异步传输中，每个字需要2.5个帧比特；而在同步传输中每个帧需要一个帧同步字，它可以涵盖几百个有用数据字。因此同步传输的效率比较高，通常被选作为快速数据传输的方法。

2.3 二元码与多元码传输

数字信号主要是由二进制符号来描述的，因此经常会误认为两者具有相同的含义。但实际上数字信号仅仅指在时间上和数值上（幅度）都是离散的信号。状态的个数（即幅度）是有限的并且是可以数得过来的，例如是2（二元码）、3（三元码）或4（四元码）。

在实际当中对信号的处理都是采用二进制的信号（见图1-48中的中间码）来进行的，但对于传输通常采用具有多个值的信号来进行。对具有4个状态值的四元信号来说，二进制信号的2Bit相当于四元信号的一个状态。四元信号状态变化的次数也只是二元信号状态变化的一半。两者信息的含量是相同的，但信号传输速率以及相应所需的带宽却减半。这也正是下面这句话的本质：多元信号是带宽和抗干扰性的互换。

根据式1-17，在从二元码向四元码转换时，在相同的时间内带宽会减半，因此信号的动态性翻倍，这也相应地要求信道的动态性翻倍，即信道有较大的信噪比。

对带宽来说，是脉冲的宽度而不是脉冲的高度起关键性的作用。但对干扰的敏感性，却与两个脉冲高度差相关。随着信号状态值的增多，高度差会变小，因此会限制多元码传输概念的使用。通过增加发送信号的功率来增加状态值之间的高度差的办法很快就会达到技术上的极限。不论如何都会增加设备的复杂性，因此在基带信号传输时人们通常采用二元码或伪三元码（物理上有3个状态，但却只表达两个逻辑状态，在接下来的小节中我们会仔细讨论）。

在干扰较小的信道中（如在微波频率范围内进行数字定向传输或在电话信道中使用高速调制解调器）进行BP传输，因效率的原因广泛使用多元码进行传输。由于已经有图1-24，所以在这里插入介绍了多元码传输。

多元码传输在存在频谱不规则分布的干扰（非白噪声）时也有很大的优点。代替

占用频谱较宽的二元码，人们在干扰较小的频率范围内使用占用带宽较窄、具有相同信息量的多元码来传输，在一定的范围内误码率较小。它遵循式（1-17），P_S 和 P_N 不是常数，而与频率有关。

每种码型中所包含的符号数叫做编码级数 M。数字信号的速率 R 以每秒的比特数（Bit/s）的方式定义。每秒钟每个信号状态变化的次数被称为调制速率 S，单位是 Baud。在二元码的情况下（而且仅仅是这种情况），1Baud = 1Bit/s，$M=2$。以四元码传输（$M=4$）时，1Baud = 2Bit/s；以八元码传输时，1Baud = 3Bit/s 等。S 也被称为波特率或符号速率，但经常被错误地用 R 来表示。对码元值为 M 的传输码，下面的关系成立：

$$R = S\log_2(M) \tag{2-1}$$

2.4　线路编码

脉冲序列在进行同步传输时，通常要进行线路编码，它的目的有两个：

1）经过线路编码后在信号中会插入足够的同步信息，以便于在接收端产生同步时钟。

2）在信号中不再包含直流分量。

这是使用同轴电缆传输所要求的，因为信号在传输路径上通常需要经过分压器的转换。此外，在长距离的传输中，还需加入放大器和转发器等，如图 1-20 所示。如果在信息信号中不包含直流分量，那么就可以使用相同的电缆和电流对放大器等进行远距离馈电。这样做比较经济（减少土木工程），因为根据不同的传输速率要求，放大器之间的距离在 1.5 ~ 10km 的范围内不等。另外数字信息在磁介质中的存储也要求消除直流分量。

另外在进行信道的直流耦合时，有时也要求在传输时消除信号中的低频分量，因为在很低的频段上很难进行整流。

图 2-4 描述了几种可用的传输线路码型，对这些码型的说明如下。这里“+”和“-”表示该码型的优缺点。注意并不是所有的码型在技术上都有明确的含义，有些只是用来进行对比说明的。

（1）单极性 NRZ 码

+：最简单的码型；

+：较小的带宽；

-：存在直流分量；

-：在出现较长的“0”串和“1”串后不能恢复时钟信号。

因此这种码型不适合同步串行传输，但在设备内部是可以使用的。

（2）双单极性 NRZ 码

同单极性 NRZ 码相比，这种码型只是试图消除直流分量。但这只是在不太容易出现的情况下才能成功，即“1”和“0”出现的概率相同才行。

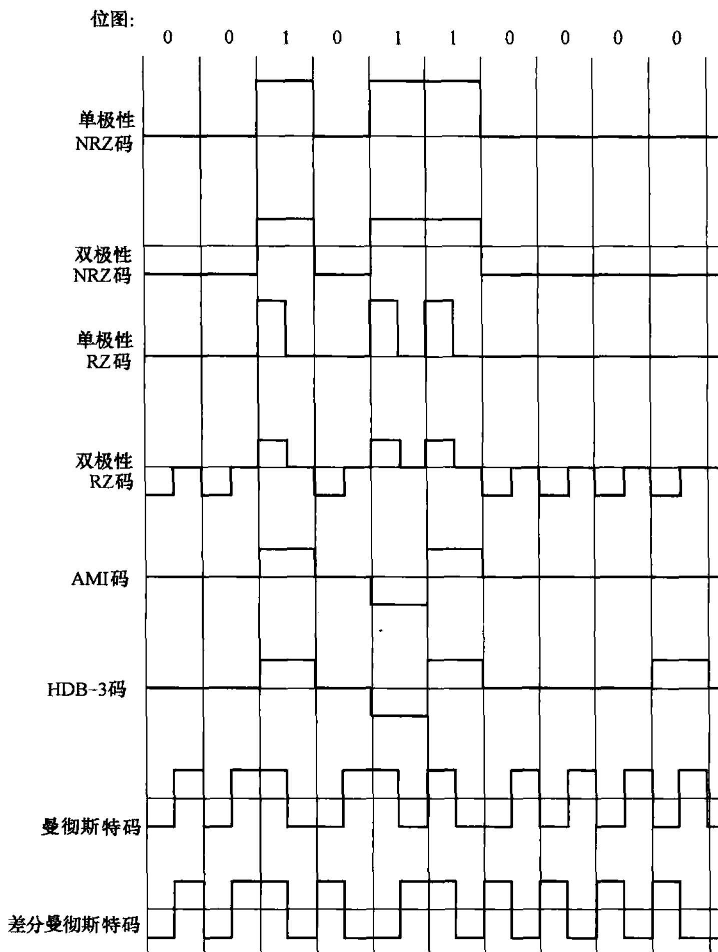


图 2-4 不同的线路码型（解释在文章中）

（3）单极性 RZ 码

+：时钟恢复是可行的（除非出现较长的“0”串）；

-：较大的带宽，因为在相同的数据传输速率下脉冲的宽度只有一半；

-：存在直流分量。

（4）双极性 RZ 码

+：时钟恢复一直是可行的；

-：需要较高的带宽；

-：涉及 3 种物理信号状态，使设备价格稍高。但逻辑上只有两个状态，因此是伪三元码。

-：只有当 0 和 1 的分布是均匀的时才能消除直流分量。

（5）AMI 码

对于逻辑“1”，有两种物理状态，即“+”和“-”可用。这两种状态起源于物

理上的翻转（如从 -5 ~ 5V），并交替进行使用。对逻辑“0”只有一种状态（即物理上的 0），因此它也是一种伪三元码。

+：需要较小的带宽；

+：没有直流分量；

-：在出现较长的“0”串后不能恢复时钟信号；

-：3 种物理状态（设备价格较高）。

（6）HDB-3 码

这种码型源于 AMI 码，但解决了出现较长的“0”串时不能恢复时钟的问题。如果在 AMI 码中出现了连续的 4 个“0”，那么就会被 000V 或 A00V 这两个组合中的一个来代替。这里 A 和 V 都代表位“1”，在 AMI 码中用正极性或负极性来表示。V-比特表示此时要打破 AMI 码的编码规则，即传输位“1”时使用与上一个比特“1”相同的极性。使用这样的规则可以将位串 0000 与 0001 区分开来。这样 AMI 码中的 1000 个连续的“0”就可以用 000V000V000V000V 等来代替。这里 V-比特破坏了 AMI 码的规则，所有的 V 都具有相同的极性。但这却产生出不希望有的直流分量。因此对 0000 有两种不同的表示方法可以选择，选择 000V 或 A00V 的目的是要使 V-比特的极性交替变换。在两个 V-比特之间一直要保持有奇数个 A-比特，而 A-比特可以从 A00V 中或是 AMI 码流中获得。使用这样的方法就不会产生任何的直流分量，见表 2-2，HDB-3 码会使用 4 种不同的组合方式来代替 AMI 码流中出现的 0000。

表 2-2　HDB-3 码对码流的修正方法

	上一个码的极性 ± 符号	
	+	-
+符号极值	-00-	000-
-符号上次编码规则的破坏	000+	+00+

在 HDB-3 码中，最多会出现 3 个连续的“0”，这也正是这个码型名字的由来。在实际中 HDB-6 码也被广泛使用。HDB-3 码最多会产生 4Bit 的延迟，这在高速率传输中不会产生任何问题。

（7）曼彻斯特码

每个脉冲只是码元宽度的一半，在码元中间进行跳变。“1”用“+”和紧随其后的“-”来传输，而“0”用“-”和紧随其后的“+”来传输。“+”和“-”传输的时间各占一半。

+：时钟恢复一直是可行的；

+：消除了直流分量；

-：需要较高的带宽。

（8）差分曼彻斯特码

在位开始时改变信号极性，表示逻辑“0”；在比特位开始时不改变信号极性，表示逻辑“1”，而转变的方向无所谓。通常在用双绞线传输时使用该码型，因为不论极性如何变换，接收端总能正确接收。其他的优缺点同正常的曼彻斯特码相同。

（9）其他码型

还有一些其他的编码方法，这里不再赘述。它们的名字如下：4B-3T码（将4个二进制码变换成3个三元码）、CMI码、条件二相码、部分响应码。

在图2-4中只有几种码型是有意义的：

1）单极性NRZ码用于设备内部的信号描述，通常与独立的同步系统并行存在。

2）HDB-3码用于同轴电缆的长距离传输，那里它需要较小的带宽是很关键的一个因素。ITU-T已经将HDB-3码定为PCM传输中2MBit/s、8MBit/s和34MBit/s接口的标准码型。

3）曼彻斯特码用于短距离串行传输，LAN就是其中的一个例子。在LAN中出于费用的考虑，人们不愿使用伪三元码。由于铜线的带宽是随线的长度而减少的，因而在LAN中曼彻斯特码的带宽需求并不成问题。人们愿意用比较便宜的双绞铜线来代替同轴电缆，因而也使用差分曼彻斯特码。

图2-5描述了不同码型在相同传输速率情况下的频谱。

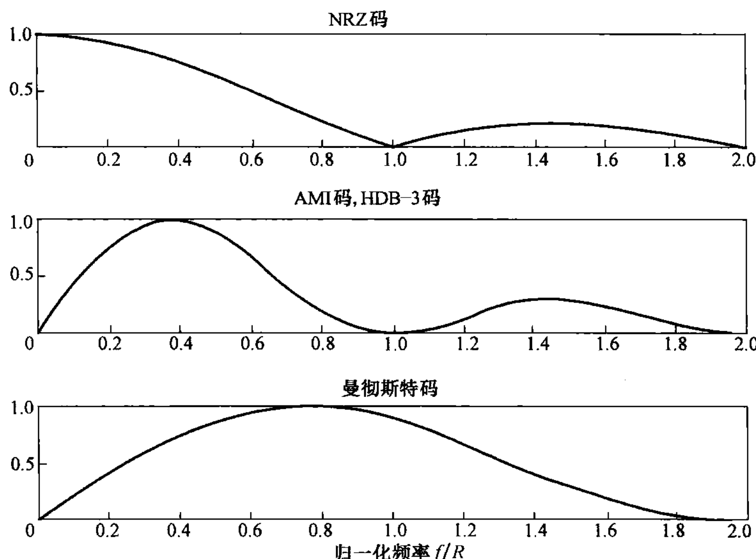


图2-5 不同码型的频谱（使用的是归一化的频率轴）

虽然上述所介绍的线路码型消除了直流分量，但传输信号中的低频成分与原来保持相同。正是这个原因，HDB-3码并不适合在电话信道进行数据传输，因为电话信道的传输频率范围是300～3400Hz，人们使用调制器去掉了低频成分，包括直流分量在内。这个例子具有代表性，可以参考图1-22～图1-24。

1）在交流耦合的BB信道传输中，线路编码有两个目的，加入时钟信息和消除直流分量。信道举例：具有输入、输出转换的同轴电缆。

2）使用TP信道（即DC耦合的BP信道）传输时，线路编码的目的只有一个，准备好时钟信息。信道举例：具有输入、输出电流耦合的铜线。

3）使用 BP 信道传输时，线路编码的目的只有一个，即准备好时钟信息。直流分量将在后续的调制器中消除。信道举例：无线传输和光传输。在光传输中，发送端的发光二极管及激光二极管起到调制器的作用。

在上述的后两种情况中，线路编码仅仅起到了插入时钟信息的作用，此时还有另外一种更好的方法能代替三元码或需要双倍带宽的码型：使用扰码器（Scrambler）。它是一个编码器，用来对数据位进行"干扰"。它的原理在加密（为了保密而进行的编码）的过程中也被应用到。扰码的结果刚好是信号中不再出现使同步变得困难的长"0"或长"1"串。

图 2-6 描述了扰码器和解扰码器的结构。在发送端将数据放入移位寄存器中，将移位寄存器进行反馈，构成一个逻辑的 EXOR 运算。门电路的作用是将长"0"或长"1"串中断。在接收端采用一个几乎相同的移位寄存器作为解扰码器。与扰码器相比，最主要的区别是解扰码器中的移位寄存器没有反馈，离开寄存器中的内容不再返回到寄存器中。解扰码器具有有限的存储器，因此在信道中出现的误码对解扰码器输出端的数据流只有一段时间的影响。解扰码器可以实现自我同步。读者最好用一个小的计算机仿真来验证扰码器的功能。

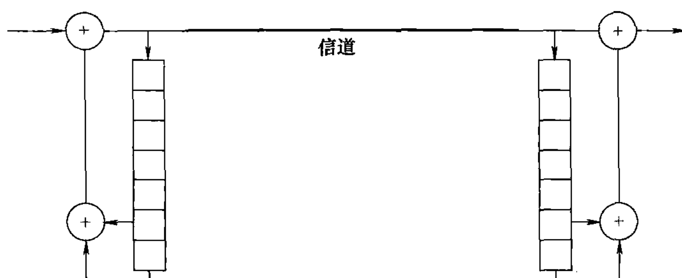


图 2-6 扰码器（左）和解扰码器（右）的门电路（ITU V. 27/V. 29）

可以使用一个扰码器（插入同步时钟信息）和一个 AMI 编码器（消除直流成分）来代替 HDB-3 码。图 2-7 描述了几种实用的线路编码器。

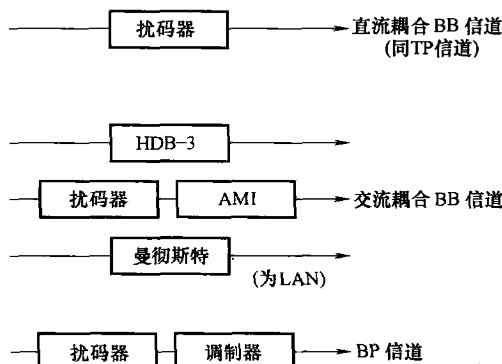


图 2-7 实用的线路编码器

2.5 传输带宽

一个数字二进制信号或一个二进制序列可以用物理上的脉冲序列来表达。为了简化起见，人们通常用直方脉冲来描述。由于信息只与脉冲的存在与否而不是与脉冲的形状有关，因此允许脉冲的波形在传输的过程中发生改变，只要不影响所包含的信息内容即可。

直方脉冲的频谱函数是 $\sin(x)/x$，比较上面的图 2-5，虽然在衰减但带宽无限。电缆及放大器具有低通特性，脉冲的边缘在传输的过程中会变圆。虽然如此，在接收端检测判决时却能够很安全地识别出原始的二进制信号值。因此使用有限带宽的信道传送包含有限信息的脉冲信号是可行的。

下面的考虑对 PAM 信号也是适用的。脉冲序列可以具有任意的高度，数字信号可以被简单地认为是量化的（例如二进制）PAM，脉冲的高度对带宽需求没有任何影响。

对最小带宽需求这个问题可以用香农（Shannon）的采样定理来回答，参见式(1-2)，对模拟信号进行采样，若要保持信息不丢失，采样频率必需大于带宽的两倍，即 $f_A > (2 \cdot B)$。若要使这个不等式的条件很好地得到满足（即过快地采样），那么采样值也是冗余的，并没有包含任何新的信息。

反过来，若以时间间隔 $T = 1/f_A$ 给出一组相互无关的数据，就可以用带宽为 $f_A/2$ 的模拟信号来表示这些数据。时间 T 相当于数字信号的步长，对二进制信号来说就是位宽度。传输二进制信号所需的最小带宽为

$$B_{ü} \geqslant \frac{1}{2T_{\mathrm{Bit}}} = \frac{R}{2} \tag{2-2}$$

在具有 4 个值的码型中，用一个码元就可以表示 2Bit。在相同的信息传输速率情况下，波特率比比特率减半。这符合通常所说的屈普夫穆勒-奈奎斯特关系及奈奎斯特带宽，即最小传输带宽为

$$B_N = \frac{S}{2} = \frac{\text{波特率}}{2} \tag{2-3}$$

2.6 码间串扰及脉冲整形

理想的模拟传输要求信号不出现任何的扭曲变形，而对无差错的数字传输，只要求在采样判决时刻信号能够被清晰地辨别出来即可。直接传输直方脉冲是不划算的，因为它占用的带宽太大，并且传输电缆和放大器本身的低通特性已经会使它发生变形。因此人们在发送端脉冲被发出之前就对它的形状进行改变，这个过程叫做脉冲整形。

但由于在频域和时域对信号进行分解之间关系并不是很明确，人们还是陷入了矛盾。矛盾的基础就是傅里叶变换的一个特性，即时间—带宽积。如果一个信号在时间轴上很尖锐，那么它在频率轴上分布得就很广，一个极端的例子是 Dirac 脉冲的频谱是一个常量；相反，如果一个信号在频域上的分布很尖锐，那么它在时间轴上扩展得

就很大，例如所有的谐波（包括 DC）。

直方脉冲的波形在时域内比较尖锐，因而在频域内占用的带宽是无限的。如果让这个脉冲经过一个低通滤波器，即让它的频率变窄，那么它在时域内就一定会变宽。因为脉冲是一个序列，这样相邻的脉冲间就会相互干扰。这种现象被称作码间串扰（Inter Symbol Interference，ISI），这样就会出现如下的问题，即两个“1”脉冲之间的“0”不能被识别出来。

但根据式（2-3），带宽有限的传输是可行的。看一个极端的例子，我们使用图 1-9 中最上面的图的脉冲串来代替中间图的脉冲串。这两个脉冲串所包含的信息内容相同。我们在信道上使用不同的时间间隔 T，不同的权重来发送 Dirac 脉冲序列。

对信道来说我们同样观察一种特殊的情况，即根据式（2-3）可以得出一个具有边界频率 $f_u = 1/2T$ 的理想低通滤波器。在信号的输出端，各脉冲响应出现叠加，并且随 T 的加大而向后推移，如图 2-8 所示。

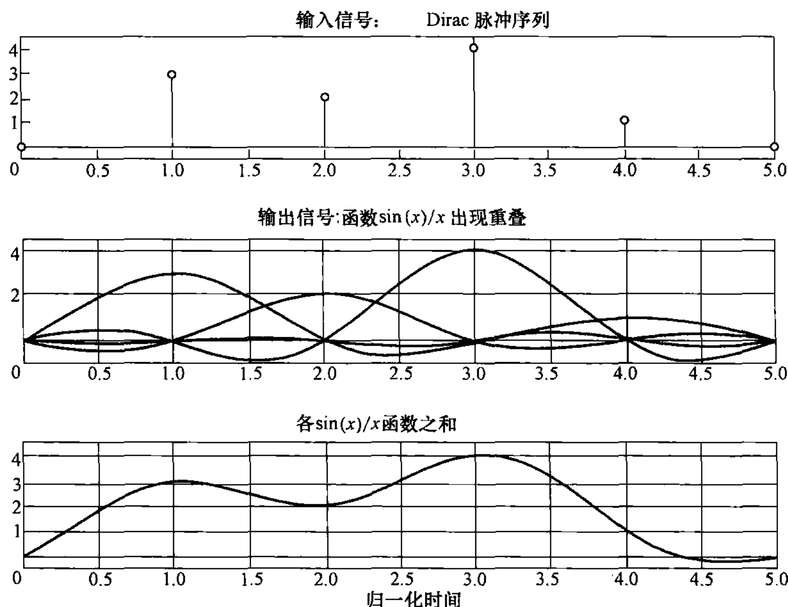


图 2-8　理想低通滤波器对狄拉克脉冲序列的响应

理想低通滤波器的脉冲响应是 $\sin(x)/x$ 函数，它在时间间隔 T 处的值为零。在图 2-8 中间的图中各脉冲响应相互叠加在一起，但在时刻 $t/T=1$、2、3、4 除外。在这些时刻，除了唯一的一个响应之外，所有其他响应的值均为零，即只有一个信号值对图 2-8 最下图所示的所有信号的叠加和有贡献。从最下面这个图中，Dirac 脉冲的权重可以无误地看出来。而在其他的时间点都有码间串扰（ISI）。上面这个例子是基于两种特殊情况的考虑，下面是一般情况。

1）在实际中，并不使用时间间隔为 T 的加权 Dirac 脉冲序列，而是使用高度不同但宽度都为 T 的脉冲。这样在频域内使用 $\sin(x)/x$ 函数会得到一个特殊的加权。这个

过程是很关键的，并且是可以补偿的。

2）理想低通滤波器是不可能实现的，因为它与脉冲响应没有因果关系。

那么现在的问题就是如果带宽边界是平滑过渡的，即同样也会导致码间串扰，是否还可以实现无差错传输。它的前提条件就是实际中的低通滤波器的脉冲响应在等时间间隔 T 上的值为零。奈奎斯特对这些问题进行了研究，并且得出了被称作奈奎斯特准则的结论：

> 奈奎斯特第一准则：脉冲经过的低通滤波器的边缘，即奈奎斯特边缘，要有一个对称点，这个对称点就位于奈奎斯特频率上。

图 2-9 描述了一种可行的滤波器与理想滤波器相比频率响应和脉冲响应的情况。两者脉冲响应的零点相同，但奈奎斯特滤波器尾部下降得更快，并且它的带宽要比理想低通滤波器的带宽略大。

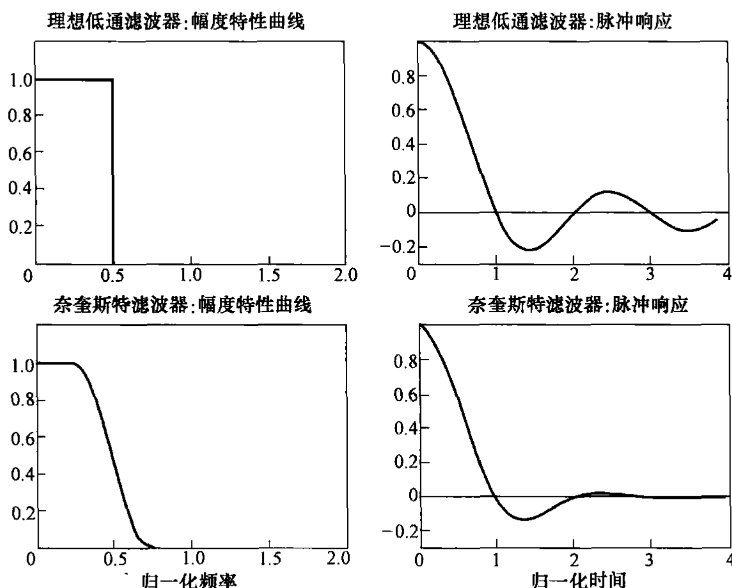


图 2-9 奈奎斯特滤波器：幅度特性曲线与脉冲响应的比较

通常使用 Raised-Cosine 滤波器来进行脉冲整形，该滤波器的幅度遵循函数为 $\cos^2(x)=0.5\cdot[1+\cos(2x)]$的曲线。在实际实现中经常是采用数字方法实现的，即经过一个插值器和横向滤波器，然后再跟一个数/模转换器。

图 2-9 中奈奎斯特滤波器的过渡部分从 $f_u=B_N-\Delta f$ 到 $f_o=B_N+\Delta f$。为此定义了所谓的滚降因子：

$$r=\frac{f_o-f_u}{f_o+f_u}=\frac{\Delta f}{B_N} \tag{2-4}$$

这里 r 是表示奈奎斯特截面陡度的一个数值，值的范围为 0（理想低通滤波器）~1

(在 $f_u = 0$ 处奈奎斯特截面已经开始出现)，通常设置成 $r = 0.5$。

由平滑滤波器截面所引起的传输带宽由式 (2-3) 的最小值 B_N (奈奎斯特带宽) 增加到如下值，即脉冲整形后所需传输带宽为

$$B_u = B_N(1 + r) \tag{2-5}$$

没经过整形的信号所占用的带宽很大，因为直方波的频谱函数 $\sin(x)/x$ 衰减得很慢。脉冲整形的目的是减少边缘非常陡峭的直方波传输所需的带宽，因此不论是基带传输还是带通传输 (参见 3.4 节) 都被使用。

设备内部具有很大的带宽可用，因为此时连接线非常短，人们一般不采用滤波器来对脉冲进行过滤。

如果脉冲整形滤波器充分遵循奈奎斯特第一准则，那么在采样点就不会发生码间干扰 (ISI)。脉冲整形滤波器并不是完全集中在发送端，而是如图 2-10 所示，由发送端的线路编码和接收端的采样器之间的所有段落共同组成。因此脉冲整形是分段构成的：

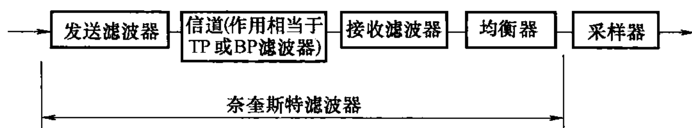


图 2-10　奈奎斯特滤波器由多个单元组成

1) 发送滤波器：由它来主要完成限制带宽的任务。在 BP 传输时，后面还紧跟着一个调制器，来实现 FDM (频分复用)，因为在这种传输中节约频谱资源是很重要的。

2) 信道：不管是否愿意，信道都会改变传输信号的频谱。有时信道的频谱特性是已知的 (例如租用线路)，有时是未知的，但至少在一个连接建立后的过程中是恒定的 (如拨号线路)，有时在一个连接存在的过程中也是变化的 (如无线链路)。信道也会产生回音，如终端混合电路 (见图 1-29) 或多径传输 (无线传输时)。

3) 接收滤波器：与相邻信道分离 (FDM 系统) 并且消除带外噪声，如图 1-67 所示。

4) 均衡器：它是整个滤波器级联组合中的最后一段，它的任务是将奈奎斯特截面进行圆滑。在实际中，均衡器和接收滤波器可以放在一起实现。

均衡器在原理上基本遵循 1.2.2 节所述内容，主要区别是在这里均衡器不必补偿信道的频率特性，而必须实现奈奎斯特第一准则，也就是说它必须将码间干扰减少到一个可以容忍的程度。有时在这里也实现一个横向滤波器 (如为适应高速数据传输)。

在图 2-8 中，在采样时刻，所发送的原始数据的幅度值绝对是可以被正确测量的。但在实际中，显然并不是那样理想，因为：

1) 接收信号与噪声叠加在一起；

2) 奈奎斯特频率并不绝对正确；

3) 恢复的时钟带有抖动。

在数字传输中幅度值是离散的，所以一定程度的噪声、码间干扰和时钟抖动不会引起问题，还是可以容忍的。

传输带宽与波特率的比值越大，越可以忍受较大的时钟抖动，对奈奎斯特第一准则的遵守程度越可以放宽。

有两种原因可以在接收端出现误码：信噪比太差或码间干扰太大，但这两种原因产生的误码还是有区别的：

> 噪声干扰产生的误码与信号内容无关，可以通过提高发送信号的功率来减少；码间干扰产生的误码与所传输的信号内容有关，提高发送信号的功率并不能减少误码。

根据式（2-1）、式（2-3）和式（2-5），可以得出可能实际应用的带宽和数据传输速率的关系，即基带传输的数据传输速率为

$$R = \log_2(M) \cdot \frac{2B_u}{1+r} \tag{2-6}$$

式中　R——数据传输速率，单位为 Bit/s；

B_u——传输带宽，单位为 Hz；

r——奈奎斯特滤波器滚降因子（值的范围在 0～1 之间，实际中常取 0.5）；

M——信号的编码级数。

传输的最大可能波特率仅与传输带宽相关，并且适用于所有的信道，包括无噪声情况下的传输。信号的编码级数可以选择任意大，因此数据传输速率可以为任意大。在有噪声的信道并且没有冗余比特（即不采用信道编码）时，

$$R(\text{数据传输速率}) = J(\text{信息速率}) \leqslant C(\text{信道容量})$$

在给定信道信噪比的情况下，可以计算出信号的最大可能编码级数。

根据式（1-17）和式（2-6），速率较大的信息传输要求带宽也较大。因此，两种表达方式“快速数据传输”和“宽带数据传输”经常被认为是相同的。这里“快速”指的是数据传输速率，并不是传播的速度。后者在无线传输中等同于光速，在有线传输中约为 2/3 光速，因此在信息网的节点中通常会引入很大的随权重递减的时延（见图 1-49）。

计算实例：有一个 48kHz 的信道，奈奎斯特滤波器滚降因子的值为 0.5，信道的信噪比为 40dB。

首先根据式（2-1）和式（2-6），波特率可计算如下：

$$S = \frac{R}{\log_2 M} \leqslant \frac{2B_u}{1+r} = 64\text{kBaud}$$

然后根据式（1-17）和式（1-18），R 的上限为

$$R \leqslant \frac{B_u}{3} \cdot SR_K = \frac{48000}{3} \cdot 40 = 640\text{kBit/s}$$

再根据式（2-1），M 的最大值为

$$\log_2 M \leqslant \frac{R}{S} \leqslant \frac{R}{S} = 10 \Rightarrow M \leqslant 1024$$

$M=1024$ 时，信道被充分地利用。最安全的是设 $M=2$，那么数据传输速率为 64kBit/s，但此时还没有考虑到信道编码的情况。

2.7 发送器、接收器和转发器

现在我们可以将到目前为止所涉及到的概念连贯起来，如图 2-11 所示。这里除了采样判决器之外，所有的模块我们都已经讨论过了。

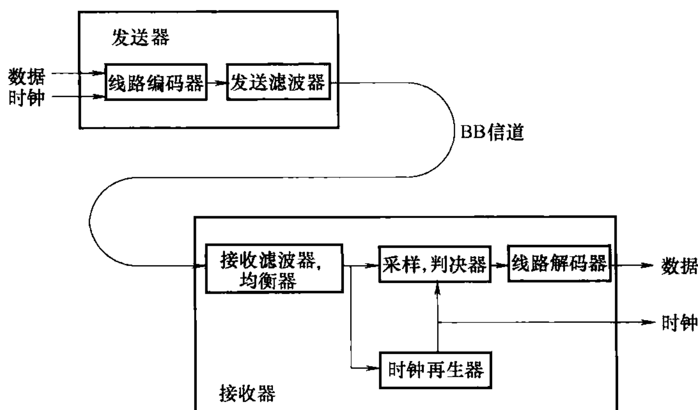


图 2-11　基带数据传输系统的组成模块

数字传输其实就是有限文字和数字构成的符号之间的交换，这个观点我们在 1.1.3 节中就已经提到了，因此在接收端需要一个判决器，来确定最可能发送的信号是什么。

在基带数字传输中，人们选择具有不同高度的但是离散的符号脉冲，因此判决器就是采样器，这错过了由时钟再生器规定的经过均衡的接收信号的时间点，并且将连续的物理信号值分给离散的逻辑信号值。对三元码或伪三元码来说，采样器需从 3 种可能性中确定一种。最简单的情况是双极性二元码传输，采样器由一个简单的比较器组成，它只需确定信号是正的还是负的即可。

可选择的信号状态越少，在接收信号具有相同的最大幅度时的判决失误就越小，如图 2-12 所示。二元码传输的抗干扰性比四元码好，后者在相同的比特率的情况下波

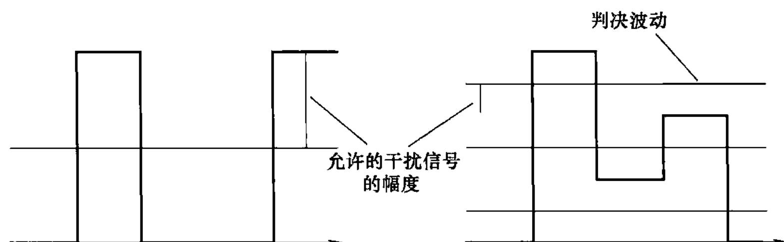


图 2-12　单极性二元码（左）和单极性四元码（右）传输时判决值加大与可允许干扰幅度

特率只有一半，因此只占用一半的带宽。

二元码传输时判决器由一个单一的比较器组成，在四元码传输时需要 3 个比较器和一个评估逻辑，每个比较器都有自己的参考标准。由于线路的衰减，这个参考标准必须从接收信号中产生，例如通过峰值检测和分压的办法。作为一种可选的办法，参考标准也可以是固定的，为此需根据输出信号的峰值检测对前置放大器进行调节。

一种特殊的比较简单的情况是采用双极性二元码传输：判决的准线一直保持在 0V，跟信号的幅度和信道的衰减都无关。经过比较器之后需要与再生时钟进行同步，图 2-13 所示为使用 D 触发器进行处理的原理图。该图描述的是一个真正的采样器。

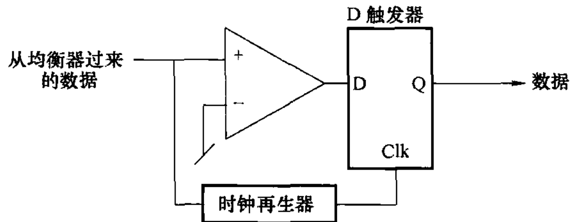


图 2-13 使用比较器，用 D 触发器进行位同步的采样器

人们越来越常用一个 A/D 转换器来代替多个比较器，如图 2-14 所示。如一个 8 位的 A/D 转换器可以提供 256 个可能的输出值。此时找到一个正确的比较参考值的问题，就转换为将 A/D 转换器的多个输出值转换成相应的多个逻辑状态。适当地使用软件，A/D 转换器也可以执行峰值检测的功能，这样有助于排序选择。A/D 转换器使用再生的时钟进行触发，可以同时起到比较器和采样器的作用。

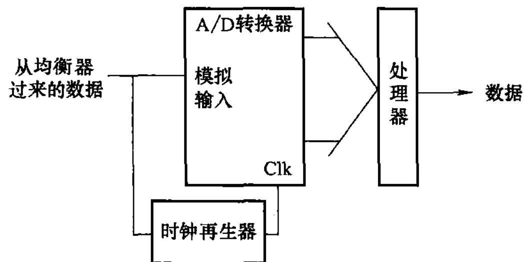


图 2-14 使用模-数转换器进行抽样和位同步

图 2-13 所示比较器的判决是很稳定的，但噪声的影响可能会使它对各个可能状态的判决有误（硬判决）。使用 A/D 转换器来判决对这种情况会有所改善。例如使用二元码传输时，在硬判决时必须将 8 位 A/D 转换器的 256 种输出信号值分成两个范围。在软判决时可以直接将采样值（代替逻辑值）送给解码器。真正的判决被取消，通过这个办法信道编码的纠错工作变得更有效，参见 4.3 节。还有一种可能是经过 A/D 转换后，需要在 3 个而不是 2 个范围内进行判决，即“1”、“0” 和 “不确定”，此时可以将该比特作为 “待确定” 位来标识。

决策任务的难易受许多因素的影响：传输带宽、滤波器的滚降因子、均衡器、时间抖动、叠加的干扰、设备的非理想性等。是否能无差错地恢复数字信号，取决于决策器输入端的输入信号及采样时间点的位置。眼图（eye pattern）能够很清楚地显示它们之间的关系。这是决策器输入端信号的示波图，这里示波器使用的是从接收器中恢复的位时钟信号。数字信号应该尽最大可能随机分布。在再生时钟不发生抖动的情况下，示波图中的信号变化相互重叠。图 2-15 最上图显示的是三元码传输的一个例子。

如果接收信号受噪声干扰，垂直方向上眼图张开的幅度就会变小，如图 2-15 中间图所示。由于再生时钟的非理想状况所导致的时钟抖动会引起眼图中信号变化的时间位置（水平方向）发生后移，所以水平方向的眼图张开幅度会变小。脉冲形状本身对眼图的宽度也会产生影响。奈奎斯特第二准则说明，在滚降因子 $r=1$ 时，水平方向上眼图张开的幅度最大[Mil97]。

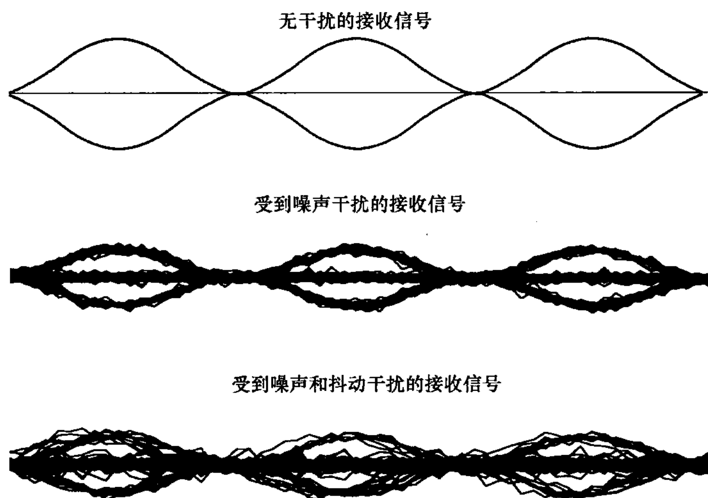


图 2-15　AMI 信号眼图的推移（在每个图中都是 80 个随机信号相互叠加）

眼图越窄，要求采样的时间越要保持准确，即对时钟抖动的容忍性越差。

可以很清楚地发现，眼图张开的幅度越大，越容易进行无差错的判决。眼图垂直张开的幅度会因下面的原因而减小：

1）ISI：均衡器工作得并不很理想，奈奎斯特第一准则受到破坏。

2）噪声：接收信号与噪声叠加在一起。

水平方向的眼图因下面的原因而减小：

1）脉冲形状：奈奎斯特第二准则受到破坏。

2）抖动：时钟再生工作得不理想。

采样判决器并没有充分利用接收信号所提供的所有可能性。图 2-16 最上图描述了没经过脉冲整形、无干扰的 NRZ 信号，这些信号将在第 1、2、3、4、5 时刻被采样。中间图描述的是相同的信号，但是被严重干扰后的情况。此时在 $t=3$ 时刻，发生了错误判决。但是我们知道在数字传输中会出现无数的符号序列，我们也知道信息就藏在这些符号中间。在 2-16 中间图中，我们期望的是一个正脉冲或负脉冲，这在时间间隔 $t=2.5\sim3.5$ 处能够找到正确的答案。判决器不令人满意的地方在于接收信号的值只能在采样时刻被判决，不能提前或拖后，因此在发送端为信号所投入的能量在接收端并没有很好地被利用。

一种改进的可能是对每个脉冲进行 3～5 次采样然后取多数值，这样可以应对时间

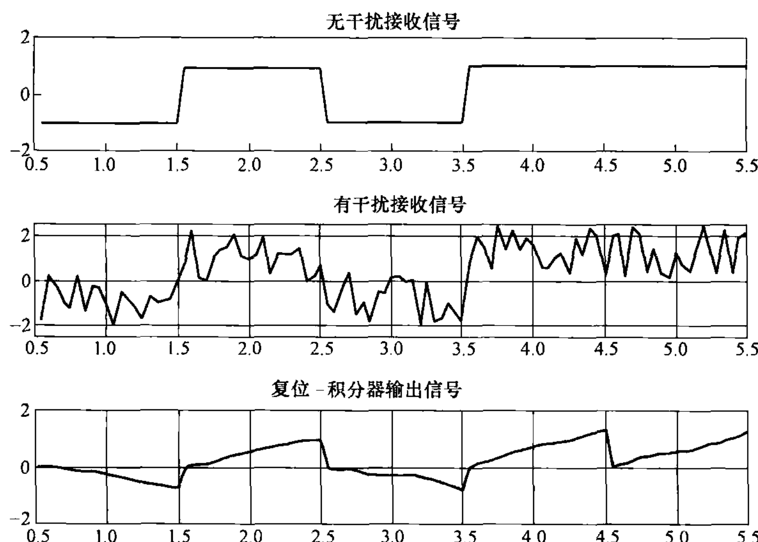


图 2-16 通过复位-积分器来检测信号

很短但峰值很大的干扰。此时判决不是在脉冲的中间而是在脉冲的边缘处就开始了。

也可以对这种方法继续扩展，对脉冲进行多次采样并对采样值进行求和。这个功能可以通过复位-积分器（Reset-Integrator）来实现，它在脉冲的开始将值置为零，在脉冲结束处输出端的值可以通过采样器来得到。图 2-16 中最下图是积分器的输出信号。在时刻 1.5、2.5、3.5、4.5 和 5.5，可以毫无问题地确定信号的正确值。

当然在实际中积分器眼图宽度是有界限的，这种方法还需细化。

上面所描述的复位-积分器只是一个高层次的原理上的抽象，即相关接收器。这个概念涉及的是采样判决器在噪声干扰很大的情况下能更好地对信号进行判决。相关接收器的来源当然要求对用统计方法对信号进行描述有较深的理解，相关内容可以参见参考文献【Mil97】、【Kam92】或【Mey02】，这里只是启发式地对它的原理进行推导。

在数字传输中接收器的任务不外乎就是猜测事先约定好的一组符号中哪个被发送过来。相关器充分利用可能的符号形状的先验知识。与抽样判决不同，相关器在脉冲存在的整个持续时间中对脉冲进行分析，而不是仅仅依据脉冲中间时刻的值对它进行判决。

香农已经证明了在信息传输的过程中能量是必需的，如式（1-24）所示。信号的能量通过对信号的二次方进行积分来获得。

$$E_{\text{Symbol}} = \int_0^T s^2(t)\,\mathrm{d}t \tag{2-7}$$

现在用接收到的信号和希望获得的信号的乘积来代替式（2-7）中的二次方值：

$$E = \int_0^T s_{\text{empf}}(t)\,s_{\text{erw}}(t)\,\mathrm{d}t \tag{2-8}$$

积分器可以通过一个滤波器来实现。通常情况，输出信号可以通过输入信号和冲激响应的卷积来计算：

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau) \cdot h(t-\tau)\mathrm{d}t \tag{2-9}$$

比较式（2-8）和式（2-9），可以计算出 $h(t)$：$h(t-\tau)$ 是冲激响应的镜像函数，相当于希望得到的脉冲。因为冲激响应是与信号相匹配的，人们也称它为匹配滤波器，它本身也叫做优化滤波器或相关滤波器。

在数字传输中，脉冲序列的持续时间是有限的，即滤波器需要的是有限的冲激响应。这也叫做 FIR 滤波器或是横向滤波器（见图 1-60），此时系数 b_i，就是希望的冲激响应的采样值[Mey02]，除非是个常数。

相关接收器需要所有可能脉冲形状的尽可能多的相关器，图 2-17 是二元码相关接收器的原理图。

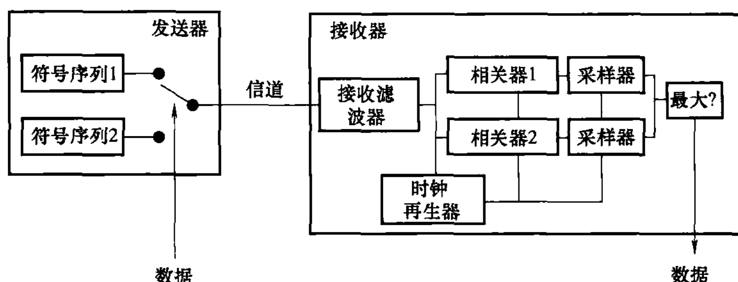


图 2-17　二元码相关接收器

图 2-17 中的符号序列相当于图 1-48 中的传输信号序列。图 2-17 不仅适用于数字基带传输，也适用于数字 BP 传输。

从 1.1.9 节中我们知道，传输信号序列的幅度和频谱都必须与信道相适应，除此之外相互之间的区别还要尽可能大。后面这点现在可以很精确地表达出来，图 2-17 中各相关器的输出必须相差很大，也就是说不同符号形状之间应该不相关，即相互正交。当式（2-10）被满足时，称信号 $s_1(t)$ 与信号 $s_2(t)$ 正交：

$$\int_{t_1}^{t_2} s_1(t) \cdot s_2(t)\mathrm{d}t = 0 \tag{2-10}$$

转发器放在发送器和接收器之间，目的是提高传输距离。转发器完整地再现数字信号并将它重新发向下一段。

转发器由图 2-11 下半部分所示的接收器后跟图 2-11 上半部分所示的发送器组成。接收器的解码器和发送器的线路编码器完成的任务刚好相反，因此在转发器中两者可以省略掉。转发器的结构图类似图 2-11 下半部分所示的接收器，但由一个发送滤波器和一个放大器来代替其中的解码器，如图 2-18 所示。

通常将远距离传输分成若干个段来进行，由转发器将各个段连接起来，如图 2-19 所示。

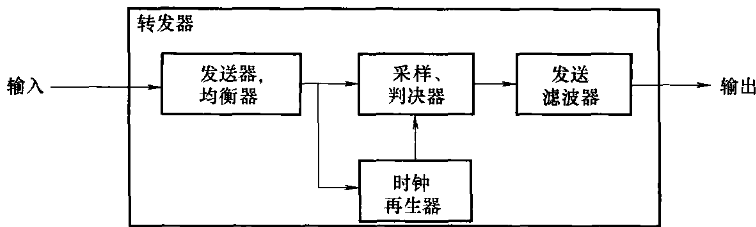


图 2-18 转发器（再生器）的结构图

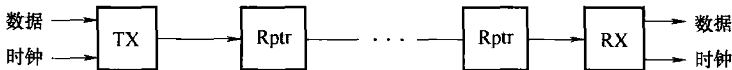


图 2-19 使用转发器的远距离传输（TX = 发送器，RX = 接收器，Rptr = 转发器）

当然双向连接是可以的，但此时设备的花费要翻倍，且通常情况下正、反两个方向都有它们各自的时钟。

问题是有可能时钟抖动从一个转发器到另一个转发器被放大了，必要的话可以使用特殊电路（虚 PLL）来减小抖动。

数字传输的健壮性由信息的免疫力，即信息在一定程度上不会被破坏来决定。如果转发器的输入信号还不是很差时，它能被 100% 地恢复，即将干扰、噪声等都去除，因此数字信号能够被传到任意远的距离。这与模拟信号不同，模拟信号虽然也可以再生，但相应的噪声值也会增加。信号的分段再生符合下面的原则：

在数字传输中增加的只是误码率，而不是干扰噪声。

第一次引入的误码当然会在下一个转发器中再生，即被保留下来。这可以通过信道编码来减少，可以在整段或者每一跳来进行保护。后者所付出的代价使转发器变得更复杂、价格更高，并且由于处理时间的增多时延会加大。数据传输速率与端到端的信道编码相比并没有增加。用 OSI 模型（参见 6.1 节）的话来说，无信道编码功能的转发器属于 OSI 模型的第 1 层，而包含道编码功能的转发器涉及第 2 层和第 3 层（也称之为网桥，参见 6.1 节）。

长距离的线路（见图 1-29 和图 2-19）出于经济上的考虑通常采用时分复用（TDM）传输方式。因为被复用的数据流中并不是所有的子流都走相同的路径，在中间会设置不同的节点（作为复用器或交换机），这样就形成了如图 1-49 所示的具有交换功能的数据网。节点的路由功能属于 OSI 模型的第 3 层功能，此时对于子数据流来说复用器的功能就相当于转发器。

一个 TDM 传输（即一个独立的物理信道）的部分失效时，会影响到无数的逻辑连接。因此在复用系统中通常一个信道预留给一个连接的运营者，而不是连接的使用者（EOW，业务信道）。特别是在双向传输中使用该方法，通过远程命令可以对发生问题的地方进行定位。转发器中也会设计有自监测功能，在出现问题时绝大多数情况下都会继续转发一个替代时钟（在转发器中自行产生的时钟），这个时钟至少可以被业务信道来使用。在发生数据丢失的地方根据错误的位置可以产生一个 AIS（Alarm Indication

Signal)，可以简单地使用事先约定好的比特位图。

现代的发送器和接收器工作在纯数字状态下。在接收器的输入端有一个快速的带有时钟的 A/D 转换器（波特率大约是图 2-14 所示的 A/D 转换器的 10～20 倍），后接一个 DSP（数字信号处理器）。使用软件它们可以完成接收过滤、均衡、时钟再生、逻辑状态判决和代码转换等所有功能，最后时钟速率可以减小到与传输率相同的值。

使用双绞线的双向传输自回音是一个很大的问题。原因是分路器非对称（见图 1-29）及电缆出现挤压（由于受到损坏及某一段落连接很差等而造成的电缆的非均匀性）。因此接收信号由远端站点的信号和自己的回音两部分混合而成，使用回音抵消电路可以减少回音。回音抵消电路实际上也是一自适应的 FIR 滤波器，它充分利用了对回音形状已知的特性，如图 2-20 所示。它根据补偿原理将干扰值进行改善，参见 1.3.1 节。

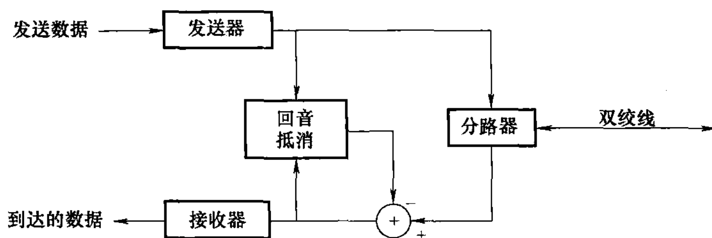


图 2-20　带有回音抵消的基带收发器（发送器-接收器）

在数据传输过程中两台机器要进行通信，将这些机器称为数据终端设备（DTE），如计算机等、和数据通信设备（DCE）、编解码器 Codec（Coder/Decoder）、调制解调器 Modem（Modulator/Demodulator）等。图 2-21 描述了从 A 到 B 进行传输时设备之间的关系顺序。在 DTE 和 DCE 之间传输可以是并行或串行的（短距离时），而在 DCE 之间通信始终是串行的。

正如刚才提到的，在双向传输中由于存在环路，可能将自己发送的信号又接收回来，也可能通过双向环路限制误码。在 ITU-T 建议 V.54 中定义了 4 个这样的环路，图 2-21 描述了用户 A 的环路。



图 2-21　根据 ITU-T 建议 V.54 用户 A 的测试环路

现代设备在运行中可以实现自检测（Built In Test Equipment，BITE），出现问题时产生环路并可以发出报警。

2.8　干扰的影响：误码

如果接收到的信号是没有噪声的，那么经过均衡之后眼图张开得刚好是最小的，

这样下去就能不断地检测到正确的状态，不会出现误码。

但在实际中所有的接收信号都叠加着噪声。在这种情况下即使通过最佳均衡，即眼图张开最大，也会出现误判。因此我们通过计算产生的误码率（Bit Error Ratio, BER），作为了解下面各项内容的基础：

1）不同传输方式的比较；

2）选择可能的用于消除误码的信道编码并确定其维数。

误码率被定义为下面的比值：

$$BER = \frac{\text{检测到的错误比特数}}{\text{传输的总的比特数}} \tag{2-11}$$

如果总的传输比特数增加，测量到的 BER 值很容易收敛到一个终结值，即误码概率 P_{Fehler}。当比特数足够大时，可以认为 $BER = P_{\text{Fehler}}$。

误码率是由于干扰产生的，它具有偶然性，因此计算误码率时需要用到概率计算知识。下面以双极性 NRZ 传输为例，概括介绍误码率的计算过程，其中难点在于计算过程中必须对干扰的特性进行假设。在对干扰的特性进行假设时，一方面结果在实际中要是可用的，另一方面计算不能太复杂。作为一个比较好的折中，多数情况下将白噪声作为干扰信号。计算得到的 BER 和通过简单测量的办法得到的 BER 会有区别的原因在于：干扰的准确模型太少，使用的设备不够理想。

白噪声的概率密度通过高斯正态分布来计算：

$$p(u_R) = \frac{1}{\sqrt{2\pi}U_R} e^{-\frac{1}{2}\left(\frac{u_R}{U_R}\right)^2} \tag{2-12}$$

式中 U_R——噪声的有效值；

u_R——U_R 的瞬时值；

$p(u_R)$——值为 u_R 的概率。

无干扰双极性 NRZ 信号在接收端有两个可能的电压值，即 $u_1(t) = +U_E$ 和 $u_0(t) = -U_E$（对应于逻辑 1 和 0）。这两个值表示信号的额定值，在采样点得到，在脉冲整形时使用的也是该值。由于通常情况下信号叠加着白噪声，因此接收信号在采样点的瞬时值为

$$u_1(t) = +U_E + u_R(t)$$

$$u_0(t) = -U_E + u_R(t) \tag{2-13}$$

$u_1(t)$ 的概率密度可以根据式（2-12）来获得，因为 U_E 的值并不是动态分布的（无干扰信号），因此被干扰信号 $u_1(t)$ 与噪声信号 $u_R(t) = u_1(t) - U_E$ 具有相同的分布，但均值是不同的，即由 U_E 代替了 0。$u_0(t)$ 的情况类似，即有如下的公式成立：

$$p(u_1) = \frac{1}{\sqrt{2\pi}U_R} e^{-\frac{1}{2}\left(\frac{u_1 - U_E}{U_R}\right)} \quad p(u_0) = \frac{1}{\sqrt{2\pi}U_R} e^{-\frac{1}{2}\left(\frac{u_0 + U_E}{U_R}\right)^2} \tag{2-14}$$

图 2-22 描述了这两个函数的曲线。

统计学中用 σ（标准差，σ^2 即方差或离差）来代替 U_R（噪声的有效值），使用该方法将时间上的平均值（有效值）与统计特性（标准差）联系起来。即无直流分量信号的功率（有效值的二次方），（或是说信号所包含的所有交流成分的功率）就等于这

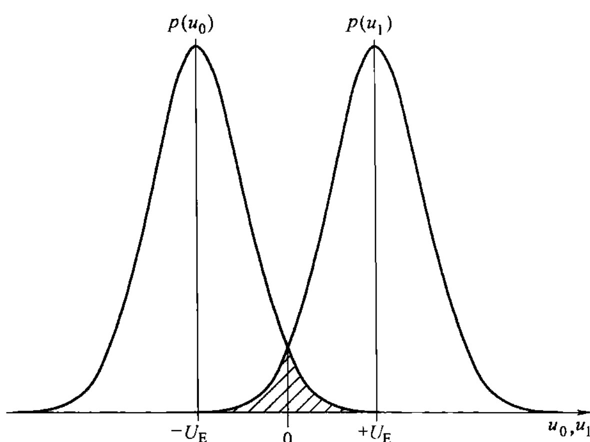


图 2-22　受到白噪声干扰的双极性 NRZ 信号的概率密度

些信号的方差。

图 2-22 中判决门限位于两个额定电压值之间，对双极性信号而言位于 0V。判决就是比较，确定接收电压的符号。危险的是图 2-22 中比较尖的部分，在这部分区域中比较器有可能会出现错误判决。

高斯噪声的概率密度以均值点对称。根据式（2-12）可以计算出 u_R 小于 U_E 的概率：

$$p(|u_R| \leqslant U_E) = \int_{-U_E}^{+U_E} p(u_R)\,du_R = 2\int_0^{+U_E} p(u_R)\,du_R \tag{2-15}$$

较大的瞬时值对正确检测信号是危险的，它的概率是

$$p(|u_R| > U_E) = 1 - 2\int_0^{+U_E} p(u_R)\,du_R \tag{2-16}$$

出现错误判决的情况是，在应该出现 1 时（额定电压 $+U_E$），刚好遇到噪声 u_R 出现很大的负值；或是该出现 0 时（额定电压 $-U_E$），噪声 u_R 刚好具有很大的正值。因此出现错误判决的的概率是式（2-16）所示的一半。将式（2-12）代入，得到：

$$p_{\text{Fehler}} = 0.5 - \frac{1}{\sqrt{2\pi}U_R}\int_0^{U_E} e^{-\frac{1}{2}\left(\frac{u_R}{U_R}\right)^2} du_R \tag{2-17}$$

上面的公式并不能得到解析解，但可以对它进行变形，得到高斯差错函数（error function）$erf(x)$：

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}\,dz \tag{2-18}$$

当然，对这个差错函数也不能进行基本的描述，但却可以对它进行数值运算。令 $x = U_E$，

$$z=\frac{u_{\mathrm{R}}}{\sqrt{2}U_{\mathrm{R}}} \tag{2-19}$$

将它代入到式（2-18）中得到

$$\begin{aligned} erf\left(\frac{U_{\mathrm{E}}}{\sqrt{2}U_{\mathrm{R}}}\right) &= \frac{2}{\sqrt{\pi}}\int_0^{U_{\mathrm{E}}} \mathrm{e}^{-\left(\frac{u_{\mathrm{R}}}{\sqrt{2}U_{\mathrm{R}}}\right)}\mathrm{d}\left(\frac{1}{\sqrt{2}U_{\mathrm{R}}}u_{\mathrm{R}}\right) \\ &= \frac{2}{\sqrt{2\pi}U_{\mathrm{R}}}\int_0^{U_{\mathrm{E}}} \mathrm{e}^{-\left(\frac{u_{\mathrm{R}}}{\sqrt{2}U_{\mathrm{R}}}\right)}\mathrm{d}u_{\mathrm{R}} \end{aligned} \tag{2-20}$$

将它代入到式（2-17）中，即可得到误码的概率：

$$\begin{aligned} p_{\mathrm{Fehler}} &= 0.5-0.5erf\left(\frac{U_{\mathrm{E}}}{\sqrt{2}\cdot U_{\mathrm{R}}}\right)=0.5erfc\left(\frac{U_{\mathrm{E}}}{\sqrt{2}U_{\mathrm{R}}}\right) \\ &= 0.5-\frac{1}{\sqrt{2\pi}U_{\mathrm{R}}}\int_0^{U_{\mathrm{E}}} \mathrm{e}^{-\frac{1}{2}\left(\frac{u_{\mathrm{R}}}{\sqrt{2}U_{\mathrm{R}}}\right)^2}\mathrm{d}u_{\mathrm{R}} \end{aligned} \tag{2-21}$$

erfc 是补余误差函数。式（2-21）将误码的概率与干扰噪声的有效值联系起来，而噪声的有效值可以通过测量技术来得到。

如果不经过脉冲整形，有用信号就是一个从 $+U_{\mathrm{E}}$ 到 $-U_{\mathrm{E}}$ 的直方波，它的有效值为 U_{E}。干扰信号的有效值为 U_{R}，上式中的比值 $U_{\mathrm{E}}/U_{\mathrm{R}}$ 正好表示的是信噪比的根（单位不是 dB）：

$$p_{\mathrm{Fehler}}=0.5erfc\left(\sqrt{\frac{P_{\mathrm{S}}}{2P_{\mathrm{N}}}}\right) \tag{2-22}$$

如果要使用 dB 来表示信噪比的值，需用下面的等式来替代式（2-21）中的商 $U_{\mathrm{E}}/U_{\mathrm{R}}$：

$$\frac{U_{\mathrm{E}}}{U_{\mathrm{R}}}=10^{\frac{\frac{S}{N}\big|_{dB}}{20}}=10^{0.05SR_{\mathrm{K}}} \tag{2-23}$$

SR_{K} 表示在 1.1.9 节中就已经介绍过的信道的信噪比，因此由式（2-22）可以得到

$$p_{\mathrm{Fehler}}=0.5erfc\left(\frac{1}{\sqrt{2}}10^{0.05SR_{\mathrm{K}}}\right) \tag{2-24}$$

图 2-23 是式（2-24）的数值分析。

上面的计算也可以这样进行：首先用一个可变值来代替 $+U_{\mathrm{E}}$ 和 $-U_{\mathrm{E}}$ 之间的平均值所对应的决策点，然后优化高度值。这样的计算基于下面的假设：

1）噪声的幅度与它的均值无关（即噪声干扰的特性与发送的是 0 还是 1 无关）；

2）1 与 0 的数目相等；

3）检测出差错所造成的损害与它是 1 还是 0 无关。

因此最优的判决高度值刚好是两个额定值的均值。这种方法也适用于单极信号和多元信号的情况，此时也可以根据直觉来选择决策点。上面 3 个假设中的最后一个与应用有关。例如在光信号设备中，由于出错信号灯变成绿色或红色是无所谓的。

根据图 2-23 可以看成在信道的信噪比为 14dB 时，误码率小于 10^{-6}，也就是说，

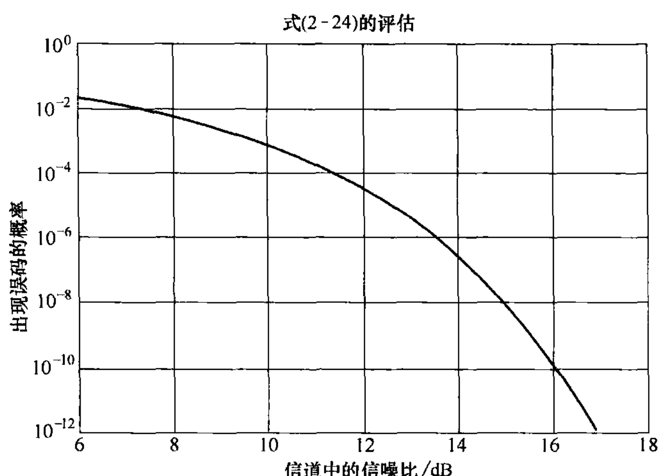


图 2-23　信道的信噪比（采样前测量）与没有经过脉冲整形的双极性 NRZ 信号误码概率之间的关系（曲线展示了理论上可能的边界值）

在传输一百万个 Bit 时平均误判一个 Bit。信道编码可以毫不费力地达到这样的误码率。比较这 14dB 的信道信噪比与表 1-7 和表 1-8 中的数据，可以看出数字传输的抗干扰能力与模拟传输的抗干扰能力相比是显而易见的。BER 为 10^{-4}的 PCM 话音信道理解起来是丝毫没有问题的。

从图 2-23 还可以看出，信噪比的一点点改善就能引起误码率较大的改善。信道的信噪比为 11.4dB 时，BER 为 10^{-4}，SR_K 为 13.6dB 时，BER 就变成 10^{-6}。信道的信噪比只提高了 2.2dB，BER 就提高了 100 倍，而将 BER 提高 100 倍相当于要将发送功率提高 1.7 倍！在实际中对一个连接而言，还会有其他附加的噪声会产生很大的影响，这无外乎就是图 1-46 介绍的膨胀效果。

猛增发送功率可以进一步降低误码率，但这起的作用并不大。比较聪明的一个办法是使用信道编码。根据维数的不同，BER 为 10^{-6}对连接不会产生任何影响，实际上对所有的差错都可以进行纠正。更聪明的办法是，在保证 BER 在一个约定的范围值内时，可以对发送功率进行调整。这点已结合图 1-46 进行了说明。

双极性二元码允许的最大干扰电压是 U_E，而在图 2-12 左侧所示的单极性二元码是 $U_E/2$。在相同的波特率下两者具有相同的数据传输速率 R，它们的区别在于传输功率：单极性码的情况下传输 0 时不消耗功率。但如果要获得相同的 BER，单极性二元码接收端的电压需要翻倍，即最大发送功率需增大 4 倍，这相当于 6dB 的区别。考虑到单极性码传输时发送可出现停顿，所以平均发送功率的区别只有 3dB，因此必须注意如何比较不同的传输方式。很遗憾不同参考文献给出的观点完全不同。

利用四元码传输时，允许的最大干扰电压继续下降，此时仍然可以通过增大发送功率的方法进行补偿。这里也存在不公平的比较：在相同的带宽下四元码的数据传输速率是双倍的。

比较公平的比较方法是采用 1.1.8 节中已经使用过的方法：作为一个基准值，

P_S/P_N并不是适用于任何数据传输速率 R 的，而 E_{Bit}/N_0 却是适用于相同的数据传输速率。下面我们观察一下没有经过信道编码而产生冗余的数据，令 $R=J$（信息速率），Bit = bit。

首先通过使用式（1-20）求出 P_S/P_N，即

$$\frac{P_S}{P_N}=\frac{P_S}{k\cdot T\cdot B}=\frac{P_S}{N_0\cdot B} \tag{2-25}$$

然后根据式（1-25）得到 $P_S=E_{Bit}\cdot R$，代入到上式中得到

$$\frac{P_S}{P_N}=\frac{E_{Bit}\cdot R}{N_0\cdot B} \tag{2-26}$$

最后还要根据式（2-3）考虑到带宽的因素。使用二元码传输时，$S=R$，因此

$$\frac{P_S}{P_N}=\frac{E_{Bit}\cdot R}{N_0\cdot S/2}=\frac{E_{Bit\cdot R}}{N_0\cdot R/2}=2\cdot\frac{E_{Bit}}{N_0} \tag{2-27}$$

P_S/P_N和 E_{bit}/N_0 之比为 2，用对数表达它们的区别刚好是 3dB。

图 2-24 所示的是误码率随对比特进行归一化后能量的变化而变化的结果。经过归一化双极性传输的曲线 b 与图 2-23 相比向左移了 3dB。x 轴选择的依据是，最左端刚好是香农极限的 -1.6dB（式（1-28））。此外曲线 c 用来与单极性码进行比较。

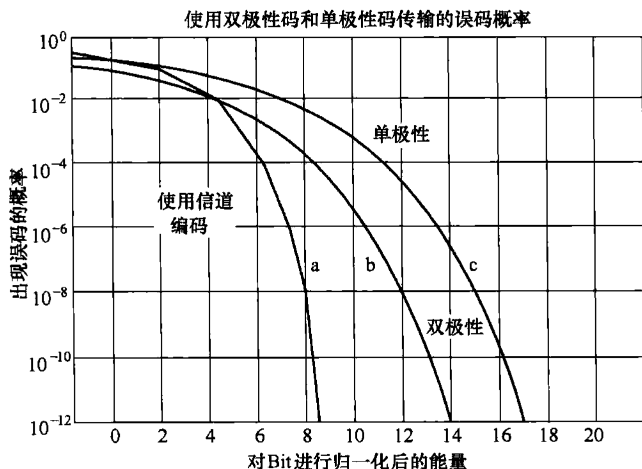


图 2-24　误码率与 SR_K 的关系，这里如参考文献【Skl88】中所述对平均功率和奈奎斯特带宽归一化

注：不同参考文献所进行的归一化处理不同，因此会影响曲线在水平方向上平移。

从图 2-24 可以看出，技术上所采用的方法会与理论上的边界相差很远。

通过增加带宽但不传输更多信息的办法，图 2-24 中双极性传输曲线可以切掉下部。减少每个比特的带宽，就可以增加传输带宽，相同时间内传输的比特数又会增加。这意味着通过传输冗余的 Bit，可以增加带宽，即 1Bit 将小于 1bit。这不外乎就是信道编码。通过目前流行的信道编码所能做到的，在图 2-24 中也大致地反映出来（曲线 a）。

图 2-24 中曲线 b 和 c 表示需要不同的传输功率，但具有相同带宽的两种传输方式；

而曲线 a 和 b 表示的是需要相同的传输功率，但具有不同带宽的两种传输方式。采用这种归一化的描述方法可以公平地对不同的传输方法进行比较。

在接收条件很差（图的最左测）的情况下，所有的曲线都收敛在 BER 为 0.5 左右。在这点上，被解码的 Bit 完全是随意的，与发送端所发出的内容完全无关。因为根据计算的前提条件，平均所发送的 1 和 0 的数目是相等的，所以会出现这样侥幸的结果，BER 仅仅是 0.5 而不是 1。

此外，BER 为 1 是绝对安全的传输，因为可以保证所有的 Bit 都与实际的相反；BER 为 0.5 意味着接收器通过监听信道和掷骰子一样随机进行判决所得到结果是相同的。

如果用相关接收器（见图 2-16）来代替采样判决器，曲线将会向左移，在相同的信噪比情况下 BER 会减小。正如前面已经提到过的，这种区别可能会变得很关键。

上面关于误码率的详细计算以及它们所产生的影响仅仅局限于 NRZ 信号和高斯噪声，但这样计算的意义有两点：首先这种结果可以定性地用于其他情况，并且会对结果的数量级有一个大致的感觉；其次通过这样做可以看出虽然理论计算是很困难的，但还是可行的。

最后，图 2-25 描述了 BER 测量的基本原理。数字随机噪声发生器（Pseudo Random Binary Noise，PRBN）作为数据源，它通过一个带有反馈环的移位寄存器来实现。在接收端还有另外一个 PRBN 发生器，它与第一个 PRBN 发生器同步。

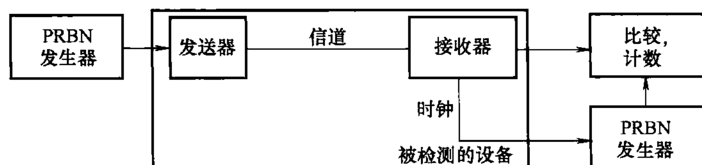


图 2-25　BER 测量的基本原理

只有当测量时间足够大时，才能通过该测量设备给出正确的误码率值。当数据传输速率为 64kBit/s（ISDN 信道）时，如果 BER 为 10^{-7}，即得到 10^4 个误码，那么必需要传输 10^{11} Bit，此时测量时间就需要 18 天！

本章到目前为止还没有提到部分响应系统。这种系统仅需要奈奎斯特带宽作为传输带宽，但却能容忍由二元码产生的伪三元码而引起的脉冲串扰。这种 ISI 是可控的，三元码也就是说它是可以计算和被补偿的。与无 ISI 系统相比，它的 BER 略有增加。有关这项技术的学习可以参考特殊的参考文献，如参考文献【Kam92】、【Mil97】及【Pro94】等。

同样，相关接收器（即匹配滤波器）技术也是一个值得深入学习的领域。

第3章 调　　制

本章的引言从1.1.9节就已经开始了。

3.1 谐波载波的模拟调制

3.1.1 引言

图3-1描述了传输段路的组成框图。其中对发送器和接收器都进行了简化处理，源信号的预处理没有放在这里。同时发送器的出口也没有加功率放大器，因为它只起到对信道衰减的补偿作用，从理论上来说对整个系统并没有多大的意义。如果接收器使用相位正确的重构的载频信号进行解调，称之为相干解调，否则称之为非相干解调。

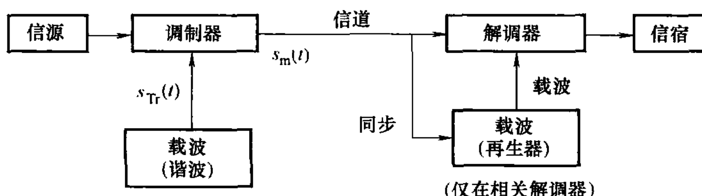


图3-1　带有调制器和解调器的传输段

在3.1节中，将下面的谐波作为载波：

$$s_{Tr}(t)=\hat{s}_{Tr}\cos[\Psi(t)]=\hat{s}_{Tr}\cos(\omega_{Tr}t+\varphi_{Tr}) \tag{3-1}$$

这里不论是幅度还是零相位都可进行调制。值得注意的是载波频率并不是可调制的，即使使用频率调制这个概念也是如此。调制后的信号具有下面的形式：

$$s_m(t)=a(t)\cos(\omega_{Tr}t+\varphi_{Tr}(t)) \tag{3-2}$$

与时间相关的幅度值$a(t)$也被称为包络线。式（3-1）中引入的参数$\Psi(t)$的导数称为瞬时频率，它包含一个固定部分（不可调制部分），即载波频率。因此频率调制正确的说法应该为瞬时频率调制。

$$\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t}=\omega_{Tr}+\frac{\mathrm{d}\varphi_{Tr}(t)}{\mathrm{d}t} \tag{3-3}$$

根据信息信号$s_{Na}(t)$所影响的参数，调制可包括如下不同的方法：

（1）只影响幅度：纯幅度调制

1）消除载波的AM（=混合）：$a(t)=ks_{Na}(t)$。

2）传统的幅度调制（AM）：$a(t)=k_1+k_2s_{Na}(t)$。

(2) 同时影响幅度和零相位

1) 单边带 (SSB) 调制。

2) 残边带 (VSB) 调制。

(3) 只影响零相位：纯相角调制

1) 频率调制 (FM)： $\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t} = \omega_{\mathrm{Tr}} + ks_{\mathrm{Na}}(t)$。

2) 相位调制 (PM)： $\Psi(t) = \omega_{\mathrm{Tr}}t + ks_{\mathrm{Na}}(t)$。

图 1-44 左半部分给出了 AM 和 FM 的信号曲线。

3.1.2 混频

混频是一种对正弦波或复杂信息信号进行频率转换的方法。频率转换的方法主要包括：分频（当前主要是通过 PLL 实现的，参见 5.4 节）、倍频（主要在微波技术中应用）和混频。

分频或倍频指输出端的频率是输入端的整分数或整数倍，随之而来信号的带宽也会相应地改变，因此该方法常用于正弦波，它是振荡器的一个组成部分。

而通过混频可以将频率转换到任意一个频段上。混频可以认为是幅度调制的一种特殊方式，也被称为线性调制。具体来说（稍后将会讨论）相当于双边带 AM 与抑制载波的混合。混频可用式 (3-2) 来描述，这里只有 $a(t)$ 受信息信号影响。

混频：

$$s_{\mathrm{DSSC}}(t) = \underbrace{ks_{\mathrm{Na}}(t)}_{a(t)}\cos(\omega_{\mathrm{Tr}}t + \varphi_{\mathrm{Tr}}) \tag{3-4}$$

因此混频相当于一个乘法器，如图 3-2 所示。

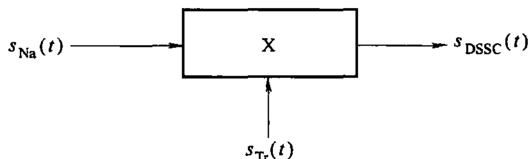


图 3-2 混频是一个纯乘法器

为了简便起见，我们首先计算两个谐波信号的乘积：

$$s_{\mathrm{Na}}(t) = a_1 A\cos(\omega_1 t + \varphi_1)$$

$$s_{\mathrm{Tr}}(t) = A_2 A\cos(\omega_2 t + \varphi_2)$$

$$s_{\mathrm{DSSC}}(t) = A_1 A_2 \cos(\omega_1 t + \varphi_1)\cos(\omega_2 t + \varphi_2)$$

$$s_{\mathrm{DSSC}}(t) = \frac{A_1 A_2}{2}\{\cos[(\omega_1 + \omega_2)t + \varphi_1 + \varphi_2] + \cos[(\omega_1 - \omega_2)t + \varphi_1 - \varphi_2]\} \tag{3-5}$$

混频时会出现和频率与差频率，混频并不是唯一的。

经过混频之后，新的频率为

$$f_{\mathrm{new}} = |f_1 \pm f_2| \tag{3-6}$$

图 3-3 描绘了上面这个例子。

通常需要对信息信号 s_{Na} 进行混频，而不仅是对上面所述的两个谐波。根据乘法的

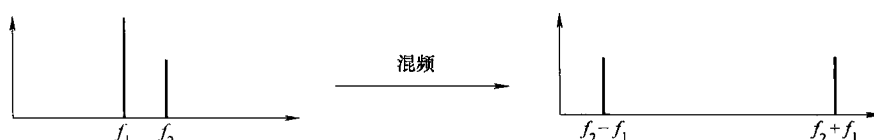


图 3-3 两个谐波信号的混频（单边频谱）

分配定律，可以先将信息信号分解成多个谐波成分（Fourier-Analyse）即采用傅里叶分析的方法，然后对每个谐波成分进行混频。

傅里叶变换是一种描述混频的好方法。因此我们先对载波进行欧拉变换：

$$s_{\mathrm{Tr}}(t)=\hat{s}_{\mathrm{Tr}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}})=\frac{\hat{s}_{\mathrm{Tr}}}{2}\mathrm{e}^{\mathrm{j}\varphi_{\mathrm{Tr}}}\mathrm{e}^{\mathrm{j}\omega_{\mathrm{Tr}}t}+\frac{\hat{s}_{\mathrm{Tr}}}{2}\mathrm{e}^{-\mathrm{j}\varphi_{\mathrm{Tr}}}\mathrm{e}^{-\mathrm{j}\omega_{\mathrm{Tr}}t}$$

然后用上面两个指数函数对信息信号 $s_{\mathrm{Na}}(t)$ 进行乘法运算。在频域内这意味着与角频率 $\pm\omega_{\mathrm{Tr}}$ 处的两个 Dirac 脉冲进行卷积。而卷积则意味着频率的平移，如图 3-4 所示。

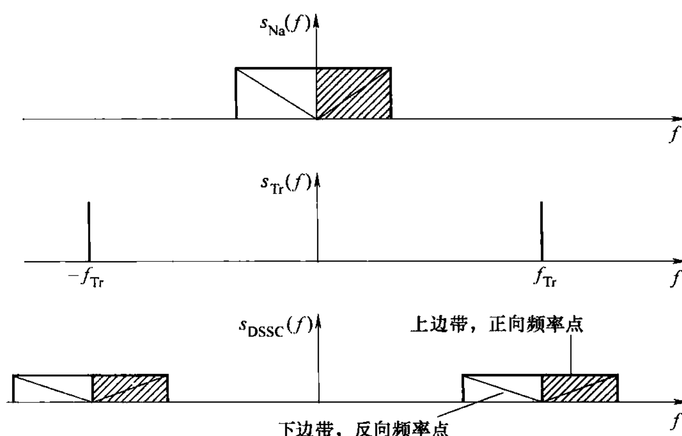


图 3-4 基带信息信号与高频正弦载波的混频（双边带频谱）

混频的不定性在于出现了两个边带。下边带就是所谓的反向频率点，因为在这里原始的信息信号的较低的频率在调制后的信号中变高了。

这里的两个边带的缩写为：

USB，LSB = 下边带，unteres Seitenband，lower sideband

OSB，USB = 上边带，oberes Seitenband，upper sideband

混频结果的带宽是原始信息信号带宽的两倍，每个边带都包含所有的原始信号信息。

混频会出现新的频率，它是一个非线性过程。但每个频谱成分对混频的贡献都是线性的，这也是线性调制这个名字的来源。

在技术上主要通过两种方法实现混频，即乘法混频和加法混频。

1. 乘法混频

根据式（3-4），应该对信息信号和载波进行相乘。这可以通过下面的方法实现：倍增器（如可控的放大器）、双栅 MOSFET 四极管（Dual Gate MOSFET）、运算放大器（由对数-加法-指数函数级联而成）和开关电路。

在实践中，可能的话通常都选用最后一种方案。相应的组成单元被称作推挽调制器、推挽混频器、环形混频器或平衡调制器。图 3-5 所示为环形混频器的原理框图。载波为双极性直方波，它的幅度大小为能完全通过二极管或被二极管完全阻断。这些二极管即构成了开关电路。如果 s_{Tr}是正的，二极管 a 和 c 就导通，反之 b 和 d 就导通。这种转换进行得很快，在两个方向一直有电流存在。环形混频器的核心是不存在磁化，因此它是 HF 可靠的，可以作为集成电路的一部分（当然需要外接线圈）。线圈是必要的，这样 HF 电压才不会短路。图 3-6 给出了调制后信号的时间曲线。

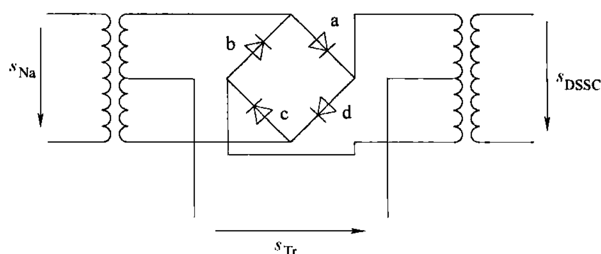


图 3-5　环形混频器

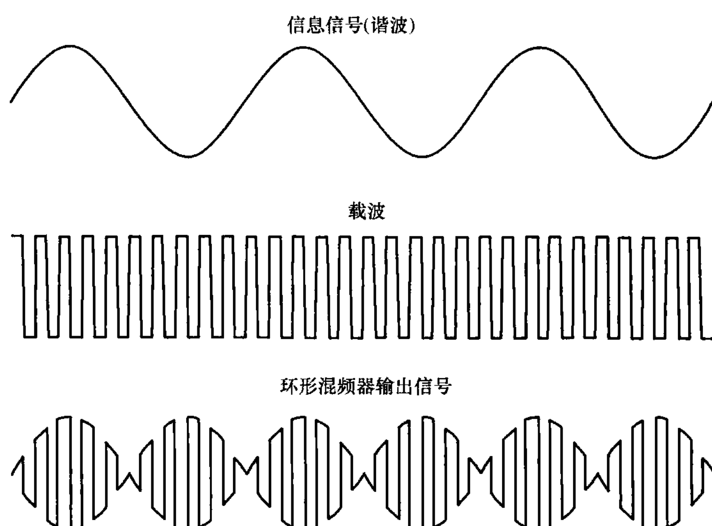


图 3-6　图 3-5 信号的时间曲线

信息信号 s_{Na} 随着载频翻转，即环形混频器的输出信号要么是 s_{Na}，要么是 $-s_{Na}$。从数学上来看，这相当于与一个幅度为 1 的双极性直方电压相乘。这种开关功能被称为 $s_R(t)$，它可以被分解成傅里叶级数：

$$s_R(t)=\frac{4}{\pi}\left[\sin(\omega_{Tr}t)+\frac{1}{3}\sin(3\omega_{Tr}t)+\frac{1}{5}\sin(5\omega_{Tr}t)+\cdots\right] \tag{3-7}$$

因此输出信号为

$$\begin{aligned}s_{DSSC}(t)&=s_{Na}(t)s_R(t)\\&=s_{Na}(t)\frac{4}{\pi}\left[\sin(\omega_{Tr}t)+\frac{1}{3}\sin(3\omega_{Tr}t)+\frac{1}{5}\sin(5\omega_{Tr}t)+\cdots\right]\end{aligned} \tag{3-8}$$

根据分配定律，式（3-8）也可以跟图 3-4 一样对每个加数项单独进行实现。它的结果是希望得到与载波的混频外加上一系列与载波频率奇数倍频率进行混频的结果。常数系数 4/π 并无大碍，因为它不含有任何信息。通常感兴趣的只是与载波频率的混频结果，其他成分都会被过滤掉。这并不会产生任何问题，因为每个单独的混频结果相互间都离得很远，图 3-7 给出了它们的频谱图。

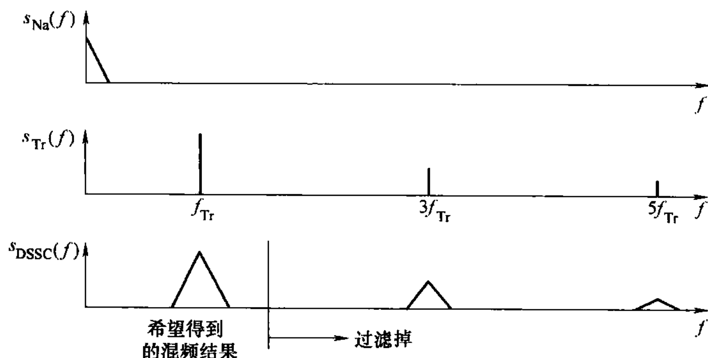


图 3-7

2. 加法混频

混频会产生新的频率，它是一个非线性过程。它可以在非线性系统中通过非线性运算如乘法，或是线性运算如加法来实现。后面这种方法被称为加法混频。

非线性系统具有非线性的曲线 $y(t)=f[x(t)]$。它可以像在 1.2.3 节中那样用幂级数来描述：

$$y(t)=a_0+a_1x(t)+a_2x^2(t)+a_3x^3(t)+\cdots$$

现在系统用两个信号的和作为激励：

$$x(t)=x_1(t)+x_2(t)$$

系统特性的二次方为

$$y_2(t)=a_2[x_1(t)+x_2(t)]^2=a_2[x_1^2(t)+x_2^2(t)+2x_1(t)x_2(t)]$$

上式中的最后一项又是一个乘积，又会产生一个和频与差频项。它的基础当然就是如图 1-65 所示的频率金字塔。

加法混频通常会充分利用二次方项。在实际中会利用二极管或晶体管的非线性特性，即 FET 的二次方特性。

3. 解调制

图 3-4 描述了调制过程的双边频谱。相似的一个想法是在接收端的解调制过程中也使用混频来实现，此时混频的频率应该与载频相等。同样又会出现和频与差频，而后者位于基带，能够表示所希望的信息，前者位于二倍载频处，将被过滤掉。下面我们用数学方法来描述这个过程：

发送端的信息信号：$s_{\mathrm{Na}}(t)$

发送端的载波信号：$s_{\mathrm{Tr}}(t)=\hat{s}_{\mathrm{Tr}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}})$

发送端出口和接收端入口的已调信号：

$$s_{\mathrm{DSSC}}(t)=s_{\mathrm{Na}}(t)s_{\mathrm{Tr}}(t)=s_{\mathrm{Na}}(t)\hat{s}_{\mathrm{Tr}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}}) \tag{3-9}$$

上述的第二个混频过程在接收端并不存在，即被调制器给分离。解调所需要的载波由接收端的振荡器产生，称为 s_{TrE}。产生一个准确的频率从技术来说很容易，因此我们设 $\omega_{\mathrm{TrE}}=\omega_{\mathrm{Tr}}$：

$$s_{\mathrm{TrE}}(t)=\hat{s}_{\mathrm{TrE}}\cos(\omega_{\mathrm{TrE}}t+\varphi_{\mathrm{TrE}})=\hat{s}_{\mathrm{TrE}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{TrE}})$$

经过混频后在解调器中有

$$\begin{aligned}
s_{\mathrm{DSSC}}(t)s_{\mathrm{TrE}}(t)&=s_{\mathrm{Na}}(t)\hat{s}_{\mathrm{Tr}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}})\hat{s}_{\mathrm{TrE}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{TrE}})\\
&=s_{\mathrm{Na}}(t)\frac{1}{2}\hat{s}_{\mathrm{Tr}}\hat{s}_{\mathrm{TrE}}[\cos(2\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}}+\varphi_{\mathrm{TrE}})+\cos(\varphi_{\mathrm{Tr}}-\varphi_{\mathrm{TrE}})]\\
&=\underbrace{s_{\mathrm{Na}}(t)\frac{1}{2}\hat{s}_{\mathrm{Tr}}\hat{s}_{\mathrm{TrE}}\cos(\varphi_{\mathrm{Tr}}-\varphi_{\mathrm{TrE}})}_{\text{希望得到的信息}}\\
&\quad\underbrace{+s_{\mathrm{Na}}(t)\frac{1}{2}\hat{s}_{\mathrm{Tr}}\hat{s}_{\mathrm{TrE}}\cos(2\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}}+\varphi_{\mathrm{TrE}})}_{\text{不希望得到的信息}}
\end{aligned}$$

不希望的混频信号将被过滤掉，解调后的信号为

$$\begin{aligned}
s_{\mathrm{Dem}}&=s_{\mathrm{Na}}(t)\frac{1}{2}\hat{s}_{\mathrm{Tr}}\hat{s}_{\mathrm{TrE}}\cos(\varphi_{\mathrm{Tr}}-\varphi_{\mathrm{TrE}})\\
&=s_{\mathrm{Na}}(t)K\cos(\varphi_{\mathrm{Tr}}-\varphi_{\mathrm{TrE}})
\end{aligned} \tag{3-10}$$

这里常数没有任何影响，但余弦函数很令人头疼。如果解调器中的振荡器与调制器中的振荡器由于偶然因素刚好相差 90°，那么经过解调之后什么也听不见。因此解调器中的振荡器必须与载波进行同步，这点在图 3-1 中已经很清楚地描述出来。一个经常使用的方法是根据式（3-9）对 DSSC 信号进行二次方运算，即

$$s_{\mathrm{DSSC}}{}^{2}(t)=[s_{\mathrm{Na}}(t)\hat{s}_{\mathrm{Tr}}\cos(\omega_{\mathrm{Tr}}t+\varphi_{\mathrm{Tr}})]^{2}$$

$$=\frac{1}{2}[s_{Na}(t)\hat{s}_{Tr}]^2[1+\cos(2\omega_{Tr}t+2\varphi_{Tr})] \tag{3-11}$$

信息信号只会影响 s_{DSSC}的幅度，而不会影响到它的频率和相位。接下来需要在两倍载频处用带通滤波器将两倍的相位提取出来。再使用一个分频器，如 PLL 或触发器就可以获得想要找的参考载波。图 3-8 所示为解调器的组成框图。二次方运算相当于与自己进行相乘，因此也可以通过混频来实现。

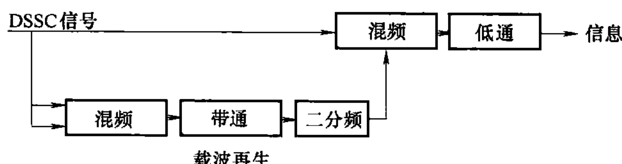


图 3-8 DSSC-解调器的组成框图

上面所有的公式推导都是使用纯乘法器作为混频器，这种纯乘法器可以通过环形混频器和带通滤波器的组合来实现。

混频及 DSSC 有很多的应用领域：频率变换和用 PSK（Phase Shift Keying，相移键控）进行数据传输。

3.1.3 幅度调制

1. 传统的 AM

载波信号也是一种谐波：

$$s_{Tr}(t)=\hat{s}_{Tr}\cos(\omega_{Tr}t+\varphi_{Tr}) \tag{3-12}$$

现在使用信息信号对它的幅度进行变换：

$$\hat{s}_{Tr}\rightarrow\hat{s}_{Tr}(t)=\hat{s}_{Tr}+s_{Na}(t) \tag{3-13}$$

式（3-13）中的常量 $\hat{s}_{Tr}$与混频有很大的区别。已调信号 $s_{AM}(t)$ 的幅度（包络）围绕着均值进行变化。

谐波信息信号 $s_{Na}(t)=\hat{s}_{Na}\cos(\omega_{Na}t+\varphi_{Na})$

调制深度 $$m=\frac{\hat{s}_{Na}}{\hat{s}_{Tr}} \tag{3-14}$$

为了简化，令 $\varphi_{Tr}=\varphi_{Na}=0$

（这些相角与频谱数量的计算无关）：

$$s_{AM}(t)=\underbrace{[\hat{s}_{Tr}+s_{Na}(t)]}_{\hat{s}_{Tr}(t)}\cos[\omega_{Tr}(t)]$$

$$=\hat{s}_{Tr}[1+m\cos(\omega_{Na}t)]\cos[\omega_{Tr}(t)]$$

$$=\hat{s}_{Tr}\left\{\cos(\omega_{Tr}t)+\frac{m}{2}\cos[(\omega_{Tr}+\omega_{Na})t]+\frac{m}{2}\cos[(\omega_{Tr}-\omega_{Na})t]\right\}$$

上述公式的中间一行表明，AM 仍然可以认为是一种混频，但此时首先给信息信号增加了一个常量。这个常量的频率为零，混频后仍然会出现与载波频率的和频与差频，

只不过此时这两项都与载波频率相等。相应地在第三行公式中，不仅出现了与混频结果一样的上边带和下边带，在载波频率 ω_{Tr}处也有成分存在。因为这个成分是由两部分组成的，所有没有像两个边带一样的系数 1/2 存在。

如果用混频的方式来描述信息信号，那么经过 AM 调制后除了载频之外还会出现两个边带。图 3-9 给出了 AM 情况下的频谱，注意与图 3-4 进行比较。

对于带宽来说，有

$$B_{AM}=2B_{Na} \tag{3-15}$$

与一般混频不同的是，载波频率会出现在调制后的信号中。如果载频成分与边带成分相比很强的话，对解调是很有用的。基于这个原因，在实际中调制深度 m 的值在 0 ~ 1 之间，即 $\hat{s}_{Tr}\geqslant\hat{s}_{Na}$。

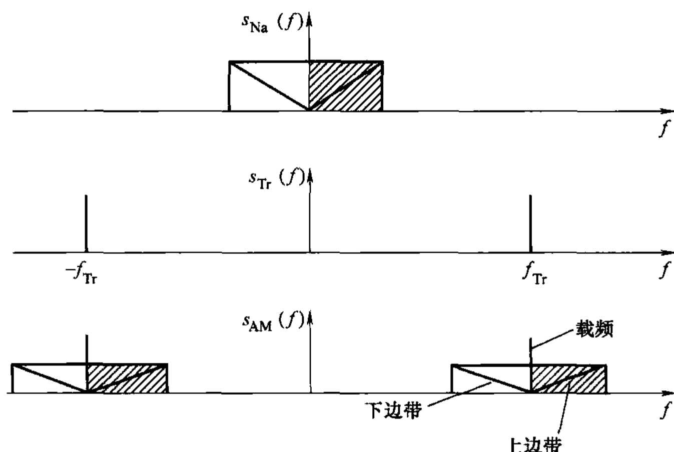


图 3-9 AM 时的频谱

图 3-10 所示为不同调制深度 m 下，经过频率为 50kHz 的载频调制后的 AM 信号。调制频率为 2kHz，载波幅度为 1V。

$m=0$ 时结果为没有调制的正常的谐波信号。载波的幅度 1V 已经是可见的。$m=0.5$ 时，幅度已经缩小为 0.5V，幅度值范围在 0.5 ~ 1.5V 之间，它始终保持在

$$\Delta\hat{s}_{AM} = \pm m\hat{s}_{Tr}$$

因此信号的包络在下面的范围内变化：

$$\hat{s}_{AM}=\hat{s}_{Tr}-m\hat{s}_{Tr}\cdots\hat{s}_{Tr}+m\hat{s}_{Tr}$$

$m=1$ 时，幅度变成 1V，它的值在 0 ~ 2V 之间变化。在实践中包络刚好接触 0 线，已经是一个极限值。

当 m 值继续增大到 1.5 时，包络的振荡继续加大。由于振荡超过了平衡位置，包络相互交叉，这种情况属于过调制，在实际中必须避免。

如果让 m 的值继续增加，幅度也随之增加，信号的功率变得没有限制。因此当信号的功率一定时，载波的幅度必需随 m 值的增加而减少。当 $m\to\infty$ 时，载波的幅度将趋于零，因此在图 3-9 所示的调制后的信号谱中就不会出现载频，而只有两个边带。

因此 $m \to \infty$ 时的 AM 就相当于混频。注意观察与 $m=1$ 时的区别，$m=1$ 时包络相互接触，混频时相互交叉。

图 3-10 的最后一幅图是倒数第二幅图的一小部分，这里包络跨过零线。可以很清楚地看到当 $t=0.125\text{ms}$ 时相位发生了改变。

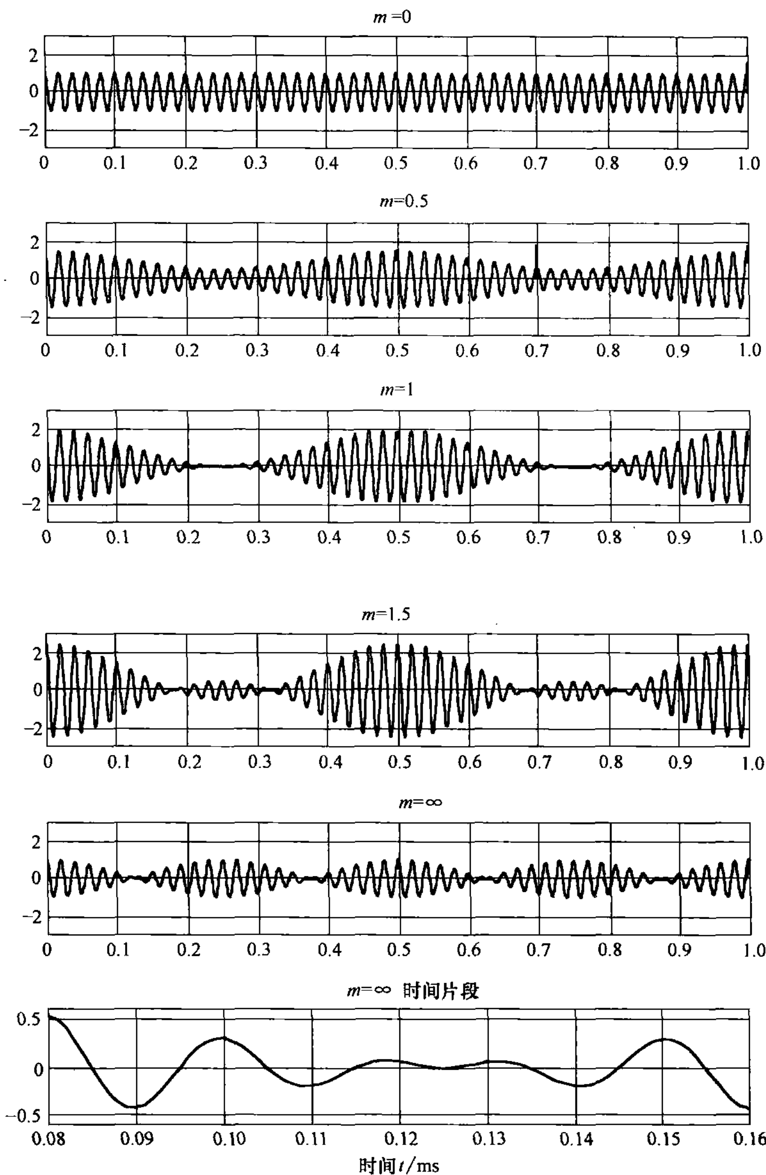


图 3-10 不同调制深度时 AM 信号的时间曲线

用矢量图来表示 AM 更清晰。我们还是观察谐波信息信号。AM 信号由 3 个矢量组成。

载波：长度：$\hat{s}_{Tr}$ 角速度：ω_{Tr}

上边线：$\frac{m}{2}\hat{s}_{Tr}$ $\omega_{Tr}+\omega_{Na}$

下边线：$\frac{m}{2}\hat{s}_{Tr}$ $\omega_{Tr}-\omega_{Na}$

由于矢量以不同的角速度旋转，结果并不是一个正弦波。矢量和在实轴上的投影为 $s_{AM}(t)$ 的瞬时值。矢量和的长度相当于包络的瞬时值。由于在 AM 情况下主要是对包络感兴趣，人们通常想象整个坐标系以角速度 $-\omega_{Tr}$旋转，这样载波矢量保持不动，而边带矢量以 $+\omega_{Na}$或 $-\omega_{Na}$的相对速度旋转。在不同的时刻画出这个矢量图，就会得到 AM 信号的包络图，如图 3-11 所示。

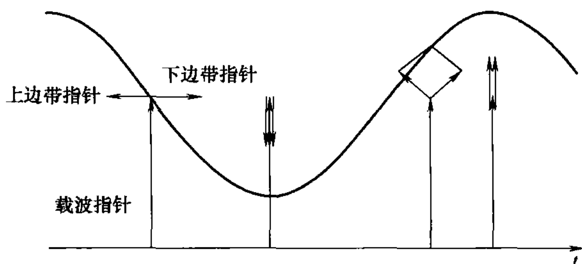


图 3-11 用 3 个矢量的和表示的 AM 信号（放大、拉伸后）的包络结构

两个边带矢量始终以载波矢量的轴为对称轴。因此 3 个矢量的和始终位于载波矢量的延长线上，即 s_{Tr}与 s_{AM}的相角始终是相同的，至少 $m<1$ 时它是成立的。

包络的最小值出现在两个边带矢量的相位与载波矢量的相位刚好相反时。$m=1$ 时，矢量和为 0，过调制（即 $m>1$）时，矢量和是负值。并且包络相互交叉，如图 3-10。通常矢量和位于载频矢量的延长线上。我们有如下结论：

传统的 $m\leqslant 1$ 时的 AM 是纯幅度调制。

传统的 $m\geqslant 1$ 时的 AM（包括 $m\to\infty$ 时的混频）是连续的 AM 与二元相位调制（相位跳跃 180°）的结合。

AM 信号的产生可以通过混频来实现，即首先给信息信号增加一个直流分量 DC，由此产生一个载波，然后将这个 AM 信号放大到希望的输出端信号功率上。

功率为几百千瓦的无线发送器需要对效率进行优化，因此会使用其他的更聪明的调制方法。详细内容可以参见参考文献【Mei92】等。

AM 的解调可以有不同的方法：相干（同步解调），如乘法解调器（混合）；非相干，如单向整流、双向整流、峰值整流。

其中最复杂的，但也是最好的方法是乘法解调器（也被称作同步解调器）。它与图 3-8 中介绍的 DSSC 解调器的方法是一模一样的，只要稍作一点变化，计算过程完全相同。式（3-10）变成如下：

$$s_{Dem}=[s_{Na}(t)+\hat{s}_{Tr}]\frac{1}{2}\hat{s}_{TrE}\cos(\varphi_{Tr}-\varphi_{TrE})$$

$$= s_{Na}(t)K_1\cos(\varphi_{Tr} - \varphi_{TrE}) + K_2\cos(\varphi_{Tr} - \varphi_{TrE}) \tag{3-16}$$

同前面一样它也需要同步。此外还产生了一个 DC 分量，由式（3-10）中的第二项构成。AM 的主要应用是传输声音信号，如长波、中波和短波的无线传输，航空飞行中 VHF 波段的无线对讲。在这种应用中信息信号不包含直流分量。图 3-8 中的解调器混频器输出端的低通滤波器被带通滤波器所替代，如图 3-12 所示。

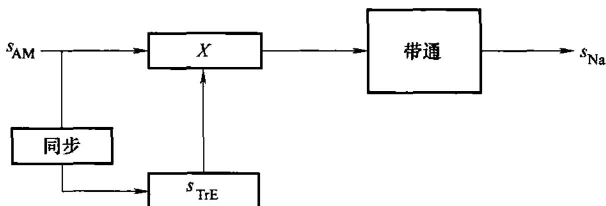


图 3-12　AM 的乘法解调器

在话音和音乐传输中，信息信号的频率从 300Hz 到几千 Hz，因此在 AM 信号载频的左右两侧出现较窄的空隙，在载波时钟再生时需使用窄带带通滤波器。

AM 的非相干解调可以使用单向或双向整流器外加带通滤波器来实现。图 3-13 的最上端是 AM 信号的时间曲线，接下来是经过单向整流和双向级整流后的信号。由包络线的形状始终可以识别出信息信号。低通滤波器可以将波动去掉，使用高通滤波器则可以将 DC 分量去掉。

整流器可以看成是带有开关电路的乘法器（混频器）。单向整流时开关函数是值为 0 和 +1 的单极性直方波，而双向整流时开关函数是值为 +1 和 −1 的双极性直方波。图 3-13 的最下端就是开关函数。整流时只要 $m<1$，这些开关函数就会自动产生。因此整流属于非相干解调类型。相反在乘法解调（相干解调）时，开关函数由一同传输过来的载波产生，因此当 $m>1$ 时（过调制）相位跳变处可以被恢复出来。

下面我们还得从数学的角度观察一下双向整流的情况。AM 信号可以描述成

$$s_{AM}(t) = [\hat{s}_{Tr} + s_{Na}(t)]\cos(\omega_{Tr}t)$$

这里我们没有考虑载波的相位，它对问题的结果不产生任何影响。双向整流时有下面的关系存在：

$$|s_{AM}(t)| = \underbrace{[\hat{s}_{Tr} + s_{Na}(t)]}_{>0,\text{若 } m<1}|\cos(\omega_{Tr}t)|$$

上式中中括号内的部分在 $m<1$ 时始终为正，余弦函数可以用傅里叶级数来展开：

$$|s_{AM}(t)| = [\hat{s}_{Tr} + s_{Na}(t)]\frac{2}{\pi}\left[1 + \frac{2}{3}\cos(2\omega_{Tr}t) - \frac{2}{15}\cos(4\omega_{Tr}t) + \cdots\right]$$

使用带通滤波器后所有超出信息信号带宽的频率将被阻止，DC 成分也会被过滤掉，剩下的仅为

$$s_{Demod}(t) = \frac{2}{\pi}s_{Na}(t)$$

AM信号

单向整流

双向整流

双向整流的开关函数

图 3-13　使用整流方法对 AM 信号进行非相干解调

下面我们来看一下最后一种 AM 解调器，包络解调器，估计它是使用最多的。如图 3-14 所示，它涉及到峰值整流。

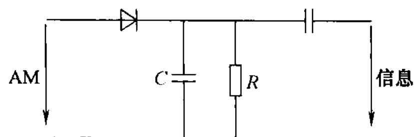


图 3-14　包络解调器（包络检测器）

AM 的电压将使用二级管来进行整流。这里电容 C 设置成可以容忍 AM 信号的最大值。超过这个值之后二级管被阻断，C 通过 R 进行放电。只要 AM 信号的电压大于 C 两端的电压，二级管就会被重新接通。放电时间常数，

即乘积 $\tau = RC$ 必需仔细进行选择。如果这个时间常数太小，输出端信号就会有过大的波动（见图 3-15 最上图）；如果时间常数太大，电容电压就会从包络中消失（见图 3-15 最下图）。C 两端的电压通过另一个电容 C 来解耦合，以消除 DC 分量。

时间常数满足下面的公式：

$$\tau \approx \frac{\sqrt{1 - m^2}}{2\pi m f_{\mathrm{Na}_{\max}}} \tag{3-17}$$

时间常数太小：太强的波纹



正确的时间常数



时间常数太大：包络失去对电容电压的控制

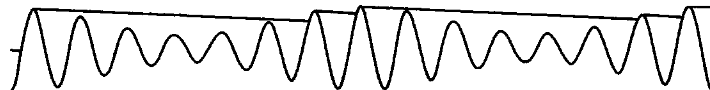


图 3-15　时间常数对包络检测的影响（实际情况下载频的频率比信息信号的频率比这里描述的大很多）

包络解调器也是一个非相干解调器。调制深度 m 必须小于 1。这种解调器的优点在于它很简单，相对来说它的输出电压也比较大。甚至可以由它来搭建一个无源无线接收器，参见 5.2 节。

后面我们还会学习另外一种解调器，这种解调器基于希尔伯特变换，非常适用于实现数字信号处理器（DSP）。

在使用无线技术之初，带有包络解调器的简单接收器是实现传统 AM 的关键性的因素。信息论研究表明，AM 不是一种很好的调制方式，因为一方面它需要过多的带宽，另一方面载波需要很多的能量投入。而载波不携带任何信息，它只能对简化解调起到作用。

但在使用无线技术的开始阶段它还是很重要的，因为要尽可能地使接收器做得简单和便宜。然而复杂度对这么少的发送器而言实在不是一个好的技术判别标准。现在，由于有源电路已经很便宜，接收器的复杂度对最终的价格影响也不是很大，因此最好不要使用 AM。但由于兼容性的问题，AM 还没有被完全放弃。

根据式（3-15）AM 需要信息信号的双倍带宽。而根据式（1-1）AM 具有一定的减少干扰的能力。现在必需解释一下这个结论。对于减少干扰而言，FM 远远优于 AM，但它需要更多的带宽。在 VHF 和 UHF 频段，由于频谱被占用得较少，还可以更充分地

利用频谱资源，因此使用 FM 而不是 AM（如 UKW 无线广播，无线手持对讲设备等）。而在长波、中波及短波范围，由于 FM 占用大量的带宽因而不被使用。

式（1-1）和式（1-19）是为噪声干扰信号而推导出的，但这对长波和短波来说刚好构不成威胁，更多的威胁来自窄带、人为产生的信号（选择性干扰）。因为这个频段很拥挤，许多发送器之间相互干扰，而这种干扰根本没有噪声信号的特点。

AM 情况下消除干扰的办法可以这样进行，理论上噪声对上边带和下边带的影响是不同的，但它的平均值在大多数情况下离原始的信息信号很近。在实际中窄带干扰对一个边带的影响会比对另一个边带的影响大很多，但对两个边带的影响还是比单独对每个只受很弱干扰的边带大。

由此可以得出这样的结论：以前由于考虑到接收器的简单性，AM 具有一定的优越性。现在只是出于对兼容性的考虑，AM 还有被使用的理由。长波、中波和短波无线通信中是这样的，但 VHF 航空无线通信中一直还使用 AM。

2. 载波抑制双边带 AM

上面我们提到了 AM 的缺点，即载波发送需要很大功率，但载波却不包含任何信息量。因此 AM 调制效率很低。这里调制效率指用于传输信息的功率与总功率之间的比值。现在我们的观点很明确了，将载波去掉。这就产生了载波抑制双边带 AM，其实它与 $m\to\infty$ 时的传统的 AM 及 DSSC 是相同的，使用复杂的接收器以较好地节省功率。使用二次方的方法重新获得载波在 3.1.2 节中就已经讨论过了。

载波抑制双边带 AM（例如 10%）也存在一定的折中，这与传统 AM 中过调制的情况是相同的。例如调制会使用一个轻微非同步的环形混频器，这会影响载波的添加物。在接收端可以使用一个滤波器将载波从接收信号中提取出来，但这需要进行处理（如使用锁相环），然后控制乘法解调器。非相干解调当然不用考虑。

3. 单边带调制

DSSC 改善了调制效率，去掉了 AM 的一个缺点。AM 另外一个缺点，即无辜占用大量带宽，可以使用单边带（SSB）调制或单边带抑制载波（Single Sideband Suppressed Carrier，SSSC）的办法来解决。它的想法是很明显的：AM 的两个边带包含有相同的信息，是冗余的，因此只传输一个边带的信息就足够了。我们进行如下的区分：

1）USB，OSB：上边带；

2）LSB，USB：下边带。

一种比较受欢迎的调制方法是滤波法，如图 3-16 所示。这里使用一个环形混频器来产生一个 DSSC 信号，并将希望的边带滤出。对这个滤波器的要求很高，为了方便可靠，它工作在固定的中间频率上。使用第二个混频器可以将 SSB 信号放在效率较高的传输频率上进行传输。第二个混频器后面的滤波器很简单，因为两个边带之间相差很远，即相差双倍的中间频率。最后的 SSB 信号经过线性放大后才被发送出去。边带的选择通过改变第一个滤波器或第一个振荡器的频率来实现。

早期 SSB 滤波器较难实现，因此很贵。后来人们不断地探索不同的 SSB 调制的方法，相位法就是其中的一种。它是一种完全补偿的方法，如图 3-17 所示。

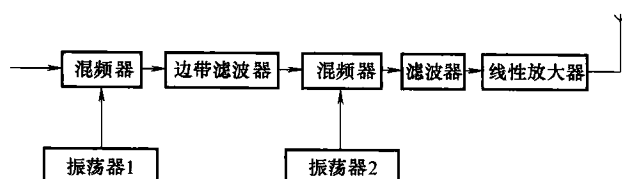


图 3-16 采用滤波法的 SSB 发送器

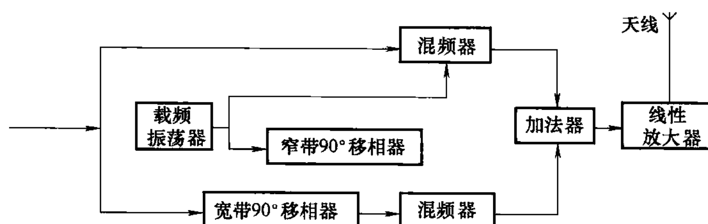


图 3-17 采用相位法的 SSB 发送器

如果交换两个移相器的输出端，可以在上边带和下边带之间进行切换。相位法涉及的问题在于 NF 移相器，因为它必须在相当宽的频带内（超过整个话音频带范围）将相位精确地旋转 90°，否则就失去了边带消除的功能。作为这种方法的一个变种，可以采用导频的方法，它只在 NF 部分与相位法有区别，并且只需使用窄带移相器，但需要两个额外的混频器和低通滤波器，因此价格很高。

相位法和导频法很少被使用，因为两者都是基于补偿的方法。使用模拟电路技术，因为漂移（温度、老化等）的存在，这样的概念很难实现；使用数字技术，可以将残留漂移问题解决掉，但它采用的完全是另一种方法。使用数字技术可以构建许多更好的宽带 90°移相器，后接全通滤波器或希尔伯特变换器。希尔伯特变换器将在 5.2.2 节中详细介绍。图 3-18 所示为数字相位法。

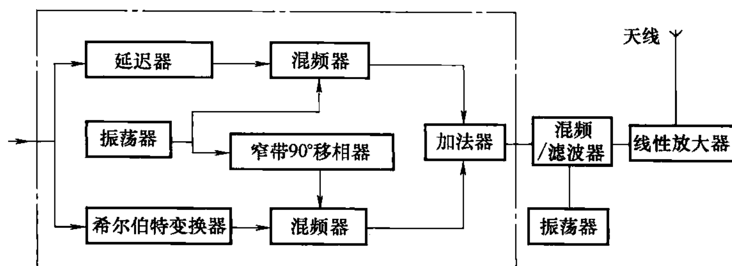


图 3-18 带有希尔伯特变换器的 SSB 发送器

图 3-18 中框起来的部分是调制器。它跟图 3-16 一样工作在固定并且较低的中间频率。之后就转换成实际的发送频率。与振荡器相连的两个混频器包含两个相差 90°的谐波，即一个正弦波一个余弦波。这采用数字技术很容易实现，相位差可以得到很好的

保证。整个调制器可以通过 DSP 中的软件实现。

解调器带有一个乘法混频器，如图 3-19 所示。与图 3-12 的区别仅在于缺少对载频振荡器的同步。这对 SSB 来说是根本不可能的，代替它的是使用敲击频率振荡器（Beat Frequency Oscillator，BFO）产生一个频率等于载波频率的谐波，并用它来控制环形混频器。缺少相位同步会产生时间偏差，但根据表 1-16，这对声音信号来说影响不大，可是对数据传输来说是不可行的，比较图 1-52。对传输音乐而言，至少 BFO 的频率要很精确。音乐是一系列谐波的组合，如频率为 500Hz、1000Hz、1500Hz 等的信号很和谐地一同发出声来。如果 BFO 只有 10Hz 的偏差，那么在解调器的传输端这些频率就变成 510Hz、1010Hz、1510Hz。它们已经不是基准频率的整数倍，发出的声音会令人无法接受。话音信号具有噪声的特点，不能破坏它的谐波顺序。定音频率（音高是固定的）也会被改变，但不会影响对它的理解，只不过是根据“蓝精灵的声音”来定音罢了。表 3-1 总结了不同信号对载波同步的要求。

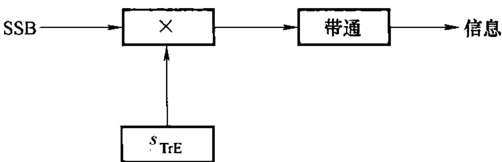


图 3-19　SSB 解调的乘法混频器

表 3-1　不同信号对载波同步的要求

	话音	音乐	图像	数据
频率同步	否	是	是	是
相位同步	否	否	是	是

第四种情况，不使用频率同步的相位同步是根本不可能的。

根据前面所述，USB 发送器传输的只是信息信号的频谱。LSB 发送器也一样，但却在反向频率点引入一个翻转。SSB 发送器可用于复杂的调制方法，如数据传输。此时可以将复杂的信号，特别是数字用 DSP 技术实现到较低的（由 DSP 的计算速度来限制）中间频率上，然后用 USB 发送器在有效的传输频率上发送。图 3-16 和图 3-18 已经说明了这个方法。

SSB 的矢量图来源于图 3-11，只是将载波矢量和一个边带矢量给消除掉，其他矢量与载波矢量相比可以是任意方向。

SSB 是连续 AM 和连续相位调制的组合

只有这样，在解调时的包络检测中才不会出现问题。这里值得提醒的是：DSSC 是连续 AM 与二元 PM 的组合。

使用 SSB 解调也可以将其他所有正常的 AM 信号进行解调。现代带有乘法解调器的远距离接收器充分利用了这点。正如前面已经提到过的，短波波段的 AM 广播受有选择的噪声的干扰，即一个边带受的干扰比一个边带大很多。ECSS（Extracted Carrier Selectable Sideband）这个概念会改善 AM 解调器：在使用乘法解调器之前用一个带选择开关的滤波器，可以有选择地对 AM 信号的上边带或下边带进行提取（即由 AM 到 SSB 转换），然后在基带进行混频。同时接收器的载波振荡器对 AM 载频进行同步，因此可以毫无障碍地收听音乐。同步电路的作用相当于对频率进行精细的调谐。

SSB 有两个优点：节省带宽和节省能量。

因此 SSM 的应用还是很广泛的：

1）短波波段内的话音广播，通常与动态压缩（音节压扩）结合起来使用，如 1.3.2 节中所描述的那样。

2）数字调制中的频率变换（数据传输）。

3）电话传输中的频分复用（FDM）。

那么 USB 和 LSB 哪个更好呢？结果很简单：两种方法同样重要。从商业的角度上来看，USB 申请了专利。对业余无线电爱好者的设备来说，工作频率在 10MHz 以下使用 LSB，10MHz 以上 USB。这有历史上的原因：20 年前这样做能让设备制造变得简单。

值得好好研究一下使用卫星的 UHF 传输：卫星的盒子上装有一个收发器，包括一个 500kHz 带宽的接收器、混频器及一个发送器。卫星将所有的可听频段内的信号放大并在另外一个频段上发回。低空卫星的对地速度较大，因此多普勒效应的影响很明显。为此经过卫星混频之后在反向频率点上还继续使用该信号。如果卫星使用 USB 信号发送，则人们收到的是 LSB 信号，用这个办法可以补偿上行链路和下行链路的多普勒效应。

4. 残边带调制

如果信息信号包含一直到 0Hz 的低频成分，则不能使用 SSB。因为这样要求边带滤波器必需做得很陡，并因此会带来很大的时延。希尔伯特变换器在这么低的频率也会出现难题。在这种情况下使用残边带（Vestigial Sideband，VSB）调制会有帮助。此时不希望的边带信号的一部分随之传输，如图 3-20 左边图所示，此后的调制过程采用滤波方法。根据载波的强度 VSB 分成全载波、部分载波及抑制载波 3 种情况。

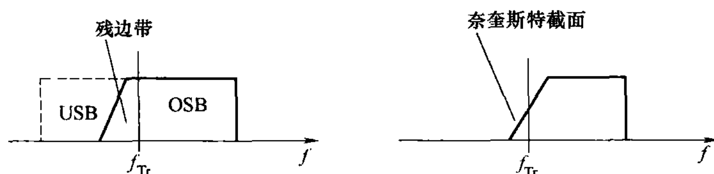


图 3-20 VSB 调制：发送器（左）和接收器（右）中的滤波

在接收端使用正常的 SSB 解调器，这样希望的边带和不希望的残留边带交互折叠在一起。为了使在这个过程中不会出现错误，在解调之前对 VSB 信号使用一个带有奈奎斯特截面（Nyquistflank）的滤波器进行滤波，如图 3-20 所示。奈奎斯特截面是一个有一个对称点（没必要是线性的）的滤波器截面。因为两个边带最初是对称的，希望得到边带的被切割掉的部分正好可以用已经传过来的不希望得到的边带部分来补偿。

在解调器之前的某个部分必需要增加奈奎斯特滤波器。这个滤波器可以完全在接收器中实现（这样不好，因为不能节省传输带宽），或是完全在发送器中实现（不好，因为被削弱的残留边带部分会被严重的干扰），或是分布在发送器和接收器中来实现。在实际中会使用第三种方法，此时发送滤波器比较陡峭（因此价格比较高），而接收器比较平缓（因此比较便宜）。

VSB 最典型的应用是电视图像传输（亮度信号），使用 VSB 进行数据传输也是可

行的。

5. 不相关边带

在ISB（不相关边带）调制中，AM信号的两个边带包含有不同的信息。ISB可以使用一个USB信号与一个LSB信号相加的方式来实现，根据需要可以在中间增加载波。

6. 正交AM

式（3-10）适用于使用乘法解调器的AM信号的解调。如果解调器中的混频器使用的振荡器与发送器中的相位相差90°，基带信号就会丢失，因此载波的同步是很重要的。正交AM在这种情况下会有很大的优势，此时两个传统的AM信号或是两个DSSC信号，使用相同的但却正交的载波相互交织，如图3-21所示。在接收器中使用两个乘法解调器，它们的载波也是正交的，并且与发送器中的载波同步。

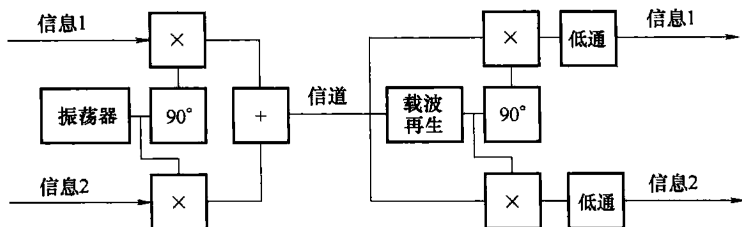


图 3-21

QAM也被用在电视技术中，即用来传输两个颜色差分信号，参见5.1.2节。它的另外一个比较重要的应用是数字带通传输。

SSB、ISB及QAM在带宽的利用以及调制效率（从功率的角度）上相差无几。

3.1.4 相角调制

相角调制（WM）工作的基础起源于1936年H. Armstrong的工作，WM指的是下面调制方法的总称：频率调制（FM）和相位调制（PM）。

FM和PM应用广泛，因此我们在一起对它进行讨论。

1. 大偏差相角调制

根据式（3-2），WM信号可以描述为

$$s_{\mathrm{WM}}(t)=\hat{s}_{\mathrm{Tr}}\cos[\Psi(t)]=\hat{s}_{\mathrm{Tr}}\cos[\omega_{\mathrm{Tr}}t+\varphi(t)] \tag{3-18}$$

它的幅度保持不变，只对幅角 $\Psi(t)$ 进行调制，如图3-22所示。载波频率 ω_{Tr} 不是调制参数。因为根据傅里叶变换，ω_{Tr} 与时间 t 有关，它不是可以自由选择的函数。

式（3-18）中余弦函数的幅角的量纲是弧度。对这个幅角求导得到的量纲是rad/s，与角频率有相同的量纲。这个导数因此也被称作瞬时（角）频率。PM与FM的区别在于载波的相位是如何起作用的：

$$\Psi(t)=\omega_{\mathrm{Tr}}t+\varphi(t)=\omega_{\mathrm{Tr}}t+\underbrace{K_{\mathrm{PM}}s_{\mathrm{Na}}(t)}_{\text{量纲:角频率}}\Rightarrow \mathrm{PM}$$

$$\underbrace{\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t}}_{\text{瞬时角频率}}=\omega_{\mathrm{Tr}}+\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t}=\omega_{\mathrm{Tr}}+\underbrace{K_{\mathrm{PM}}s_{\mathrm{Na}}(t)}_{\text{量纲:频率}}\Rightarrow \mathrm{FM} \tag{3-19}$$

因此 FM 应该称作瞬时频率调制。图 3-23 给出了 FM 与 PM 之间的区别。

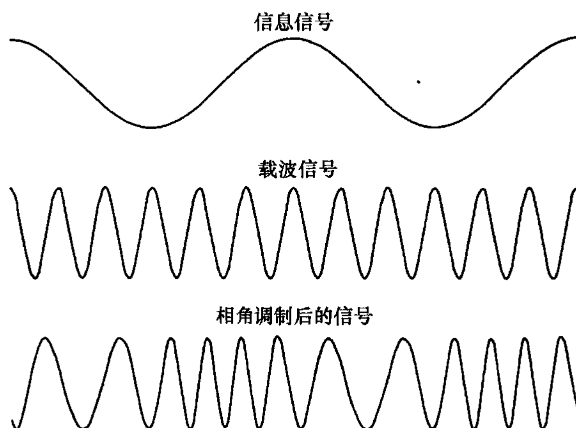


图 3-22　相角调制中的不同信号

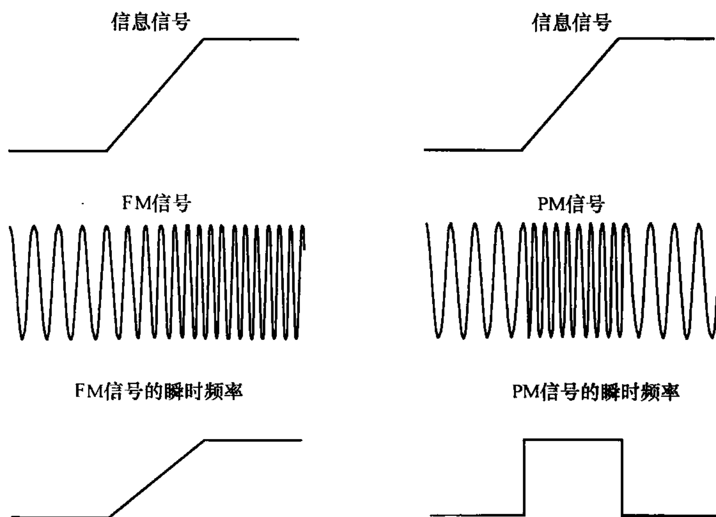


图 3-23　FM 与 PM 之间的区别

作为信息信号我们仍先观察谐波信号：

$$s_{Na}(t)=\hat{s}_{Na}\cos(\omega_{Na}t+\varphi_{Na})$$

因此，PM 信号的幅角为

$$\begin{aligned}\Psi_{PM}(t)&=\omega_{Tr}t+K_{PM}\hat{s}_{Na}\cos(\omega_{Na}t+\varphi_{Na})\\&=\omega_{Tr}t+\underbrace{\Delta\varphi_{Tr}}_{\text{相位偏差}}\cos(\omega_{Na}t+\varphi_{Na})\end{aligned}\tag{3-20}$$

FM 信号幅角的导数为

$$\frac{\mathrm{d}\Psi_{\mathrm{PM}}(t)}{\mathrm{d}t} = \omega_{\mathrm{Tr}} + K_{\mathrm{PM}}\hat{s}_{\mathrm{Na}}\cos(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}})$$

$$= \omega_{\mathrm{Tr}} + \underbrace{\Delta\varphi_{\mathrm{Tr}}}_{\text{频率偏差}}\cos(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}}) \tag{3-21}$$

因此 FM 信号的幅角为

$$\Psi_{\mathrm{FM}}(t) = \int[\omega_{\mathrm{Tr}} + \Delta\omega_{\mathrm{Tr}}\cos(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}})]\mathrm{d}t$$

$$= \omega_{\mathrm{Tr}}t + \frac{\Delta\omega_{\mathrm{Tr}}}{\omega_{\mathrm{Na}}}\sin(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}}) + \underbrace{\varphi_{\mathrm{Tr}}}_{\text{积分常数}}$$

$$= \omega_{\mathrm{Tr}}t + \underbrace{\mu}_{\text{调制指数}}\sin(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}}) + \underbrace{\varphi_{\mathrm{Tr}}}_{\text{积分常数}} \tag{3-22}$$

调制后信号的时间函数为

$$s_{\mathrm{PM}}(t) = \hat{s}_{\mathrm{Tr}}\cos[\omega_{\mathrm{Tr}}t + \Delta\varphi_{\mathrm{Tr}}\cos(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}})] \tag{3-23}$$

$$s_{\mathrm{FM}}(t) = \hat{s}_{\mathrm{Tr}}\cos[\omega_{\mathrm{Tr}}t + \varphi_{\mathrm{Tr}} + \mu\sin(\omega_{\mathrm{Na}}t + \varphi_{\mathrm{Na}})] \tag{3-24}$$

结论：如果信息信号是谐波，FM 和 PM 没有区别。

如果在调制之前对（任意的）信息信号求积分，那么 PM 会变成 FM，如图 3-24 所示。相反，如果首先对信息信号进行微分，那么 FM 会变成 PM，如图 3-25 所示。

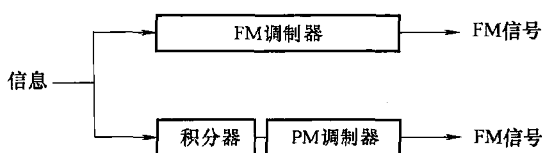


图 3-24 使用积分器和 PM 调制器的 FM 调制

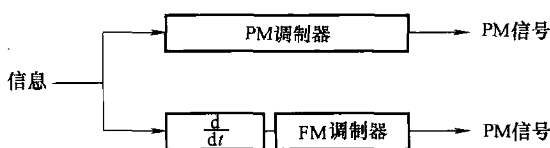


图 3-25 使用微分器和 FM 调制器的 PM 调制

图 3-26 是 WM 的矢量图。同样，这里坐标系以角速度 $-\omega_{\mathrm{Tr}}$旋转。矢量的长度是固定值，它的静止状态与 φ_{Tr}有关。WM 的矢量图就像汽车挡风玻璃的雨刷。根据矢量

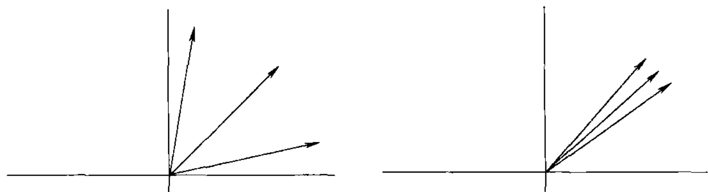


图 3-26 大偏差 WM（左）与小偏差 WM（右）的矢量图

之间偏差的大小，人们常说大偏差 WM 或小偏差 WM。

对 WM 信号的频谱不能进行通用的计算，下面我们计算一下当信息信号是余弦形式时 FM 信号的频谱。虽然这里不能进行简单的替代，计算结果也不能进行一般化，但是这个例子对一个传输信道的带宽需求还是能够得出结论的。

根据式（3-24）FM 信号可以描述为

$$s_{FM}(t)=\hat{s}_{Tr}\cos[\omega_{Tr}t+\varphi_{Tr}+\mu\sin(\omega_{Na}t+\varphi_{Na})]$$

零相角 φ_{Tr} 和 φ_{Na} 与零时间点的选择有关，与幅度谱没有关系，因此可以设成 0，即：

$$s_{FM}(t)=\hat{s}_{Tr}\cos[\omega_{Tr}t+\mu\sin(\omega_{Na}t)]$$

FM 信号可以用贝塞尔级数（也被称为柱面函数）来表示，即

$$\cos[a+x\sin(b)]=\sum_{n=-\infty}^{\infty}J_n(x)\cos(a+nb)$$

式中　$J_n(x)$——阶数为 n 的第一类贝塞尔函数在 x 处的值。

不同贝塞尔函数的值可以用图表来表示。图 3-28 以图形的方式描述了贝塞尔函数的值。将它们代入到 FM 信号中有

$$s_{FM}(t)=\hat{s}_{Tr}\sum_{n=-\infty}^{\infty}J_n(\mu)\cos(\omega_{Tr}t+n\omega_{Na}t)$$

可以对上面这个级数一项一项地进行傅里叶变换，这里 $J_n(x)$ 只是一个常数系数。利用下面的近似：

$$\cos(\omega_0 t)\circ\!\!-\!\!\bullet\,\pi[\delta(\omega-\omega_0)+\delta(\omega+\omega_0)]$$

得到下面的频谱函数：

$$S_{FM}(\omega)=\hat{s}_{Tr}\sum_{n=-\infty}^{\infty}J_n(\mu)\pi[\delta(\omega-\omega_{Tr}-n\omega_{Na})+\delta(\omega+\omega_{Tr}+n\omega_{Na})] \quad (3\text{-}25)$$

理论上 FM 的频谱是无限宽的。 当信息信号是谐波时，它的频谱由多条线组成，这些线以载波频率对称，并且线之间的间距为 ω_{Na}。

图 3-27 给出了这样的频谱的一个实例。

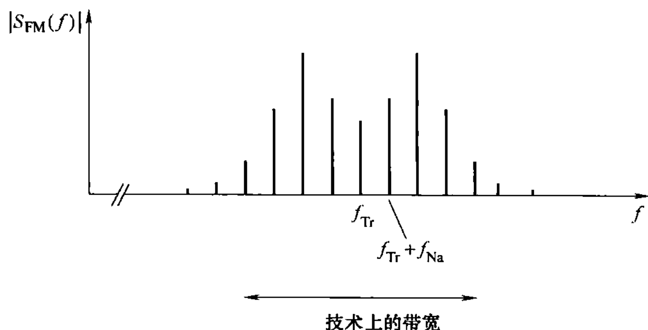


图 3-27　信息信号为谐波，$\mu=3$ 时 FM 信号的频谱值

每条线段的值可以用贝塞尔函数来确定。图 3-28 给出了正数 n 的贝塞尔函数值。n 为负数时的值可以用下面的等式来推导：

$$J_{-n}(\mu)=(-1)^n J_n(\mu)$$

即 $J_{-1}(\mu)=-J_1(\mu)$，$J_{-2}(\mu)=J_2(\mu)$ 等。从图 3-28 中可以，根据选定的调制指数在相应的垂直轴上可以得到相应的频谱值。

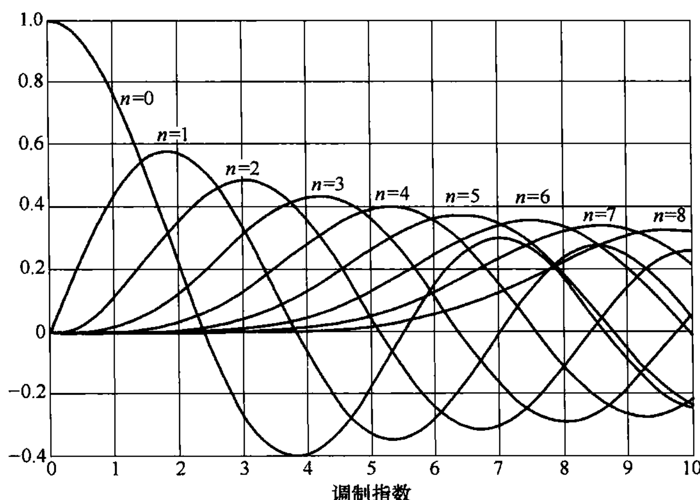


图 3-28 第一类贝塞尔函数，阶数为 0 ~ 8

高阶贝塞尔函数比低阶贝塞尔函数减弱得慢。具体来说，在 $\mu=2$ 时，$J_4(2)$ 的值还很低，可以被忽略。同样，当 $\mu=6$ 时，成分 $J_8(6)$ 的值也可以被忽略。一般来说，式（3-25）中必需要一直考虑到成分 $n=\mu+1$ 的影响。因此在实际中 FM 的带宽是有限的。根据式（3-22），调制指数可以使用另外一种方法来描述，此时 FM 的带宽可用两种方法给出：

FM 的传输带宽： $$B_{FM}=2(\mu+1)B_{Na}=2(\Delta f_{Tr}+B_{Na}) \tag{3-26}$$

这个带宽也被称作卡森（Carson）带宽。FM 的带宽可以通过下面的方法得到：

1）瞬时频率变化的大小（这就是频率偏移），它与信息信号的幅度有关。

2）瞬时频率变化的快慢，它与信息信号的频率有关。

在 UKM 广播时最大的调制频率是 15kHz，最大的频率偏移是 75kHz。从技术的角度上来说，调制后信号的带宽是 180kHz。由于对式（3-25）中高阶成分的忽略而被修整的频谱影响非线性偏差。在对 FM 系统有很大的线性需求时，带宽要比式（3-26）有所提高。使用式（3-26），失真系数小于 1%。

在传输带宽之内，每个谱线由于贝塞尔函数的振荡有完全不同的值。谱线长度受调制指数的影响，也与信息信号的幅度和频率有关。载频谱线可能完全消失，例如当 $\mu\approx2.4$ 或 5.5 或 8.7 时。

随着调制指数的增加，μ 也会增加传输带宽。这会产生一个调制增益，即对干扰的抵抗力会增加。带宽的放大量 β 的值为

$$\beta_{\mathrm{FM}}=\frac{B_{\mathrm{FM}}}{B_{\mathrm{Na}}}=2(\mu+1)=2\left(\frac{\Delta f_{\mathrm{Tr}}}{f_{\mathrm{Na}}}+1\right) \tag{3-27}$$

这意味着，在较低频率带宽的放大值比在高频处大。因此使用较低调制频率对干扰的保护比使用高的调制频率好。观察一下接收器中解调后的信号，干扰信号的功率是非均匀分布的。高频部分的干扰较强，这正好符合图 1-68 所说的情况。所采用的补救措施同那里描述的也相同：在发送器中对信号调制之前，将高频成分加强，即将偏差增大（预加重阶段），而在接收端解调之后再将这些成分削弱（去加重阶段）。预加重阶段采用微分来实现，但是将频率在感兴趣的频带内（如对 UKW 广播来说是 15kHz）加强是没有意义的，因此使用如图 3-29 所示的电路来进行。增益响应在 3.18kHz（美国标准是 2.2kHz）处增加一个向上的弯曲，这相当于一个时间常数$R_1C=50\mu s$（美国是 75μs）。去加重阶段采用一阶低通滤波器即可。

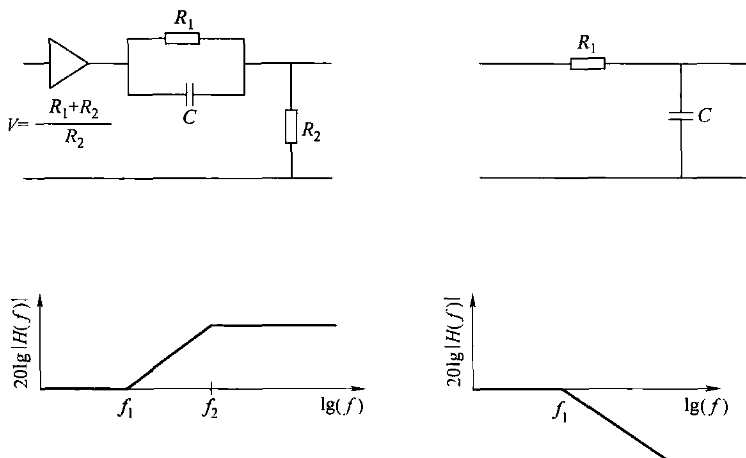


图 3-29 FM 预加重（左）和去预加重（右）的电路和频率响应特性

预加重阶段的作用是较高的频率成分采用 PM 进行传输，而较低的频率成分采用 FM，图 3-25 下半部分描述的就是这种方法。由于预加重阶段的存在，调制指数μ在实际中是一个常数，即与调制频率无关。

因为对谐波调制时（FM 情况下也是计算的这种情况下的频谱）PM 和 FM 没有区别，所以 PM 情况下的带宽可以采用类似的方法进行计算。根据式（3-20）及式（3-22），用φ_{Tr}代替μ得：

PM 的传输带宽：

$$B_{\mathrm{PM}}=2(\Delta\varphi_{\mathrm{Tr}}+1)B_{\mathrm{Na}} \tag{3-28}$$

PM 和 FM 带宽的重要区别在于

$$B_{\mathrm{FM}}=2(\mu+1)B_{\mathrm{Na}}=2(\Delta f_{\mathrm{Tr}}+B_{\mathrm{Na}})=2(K_{\mathrm{FM}}\hat{s}_{\mathrm{Na}}+B_{\mathrm{Na}})$$

$$B_{\mathrm{PM}}=2(\Delta\varphi_{\mathrm{Tr}}+1)B_{\mathrm{Na}}=2(K_{\mathrm{PM}}\hat{s}_{\mathrm{Na}}+1)B_{\mathrm{Na}}$$

为了得到一定的调制增益，频率和相位偏差的选择要远大于信息信号的带宽，所以存在下面的关系：

$$B_{\mathrm{FM}}\approx 2K_{\mathrm{FM}}\hat{s}_{\mathrm{Na}}\quad B_{\mathrm{PM}}\approx 2K_{\mathrm{PM}}\hat{s}_{\mathrm{Na}}B_{\mathrm{Na}}$$

FM情况下调制后信号的带宽几乎与调制信号的带宽无关。相反在PM情况下，在实际中信息信号的幅度和频率需为常数，这样调制后信号的带宽变化才不会太强烈。对音频和图像信号来说这种前提假设通常情况是不成立的。正是如此，UKW广播时采用的是FM而不是PM，因此预留的信道带宽能够得到更好的利用。

WM信号的功率与AM不同，与信息信号无关，值为

$$P_{WM} = \frac{\hat{s}_{Tr}^2}{2} \tag{3-29}$$

若根据式（3-26）及式（3-28）对传输带宽进行简化，那么式（3-29）中值的99%仍然集中在这个范围。

FM和PM对信道适配都是可用的，因为两者都：

1）可以将频率变换到任意频段上。

2）允许频率复用。

3）改变了带宽和对干扰的敏感性（对信息进行了重新塑造）。

FM与AM相比的优点在于：

1）发送端能量利用得更好（没有信息丢失，但载波能吸收一定的能量）。

2）可以通过设置偏差的大小，有选择性地消除干扰。

3）对非线性信道有一定的免疫力，因为此时信息会隐藏在零交叉点。因此它也可以应用于具有高有效系数的非线性功率发送等级。

FM与AM相比的缺点在于：

1）FM需要较多的带宽。这是在长波、中波及短波中不能使用FM的最关键的一个因素。

2）接收器电路的复杂性较高。在集成电路时代，这已经不是一个很大的问题了。但是VHF频段的空中飞行广播是使用AM的，如果现在要将它们改成FM，系统代价太高了。

2. 小偏差FM

小偏差FM这种特殊情况指的是调制指数的值很小，即$\mu \ll 1$。根据式（3-26）可以得到此时的带宽为

$$B_{KFM} = 2B_{Na} = B_{AM}$$

这也可以根据式（3-24）来进行计算，同样，这里也不考虑零相位：

$$s_{FM}(t) = \hat{s}_{Tr}\cos[\underbrace{\omega_{Tr}t}_{\alpha} + \underbrace{\mu\sin(\omega_{Na}t)}_{\beta}]$$

根据加法定律：

$$\cos(\alpha+\beta) = \cos(\alpha)\text{cso}(\beta) - \sin(\alpha)\sin(\beta)$$

因此，当$\mu \ll 1$时，

$$s_{KFM}(t) = \hat{s}_{Tr}\{\cos(\omega_{Tr}t)\underbrace{\cos[\mu\sin(\omega_{Na}t)]}_{\approx 1} - \sin(\omega_{Tr}t)\sin[\underbrace{\mu\sin(\omega_{Na}t)}_{\approx \mu\sin(\omega_{Na}t)}]\}$$

$$= \hat{s}_{Tr}[\cos(\omega_{Tr}t) - \sin(\omega_{Tr}t)\mu\sin(\omega_{Na}t)]$$

对上式利用下面的公式：

$$\sin(\alpha)\sin(\beta) = -\frac{1}{2}\cos(\alpha+\beta) + \frac{1}{2}\cos(\alpha-\beta)$$

因此可以得到小偏差 FM 信号时：

$$s_{KFM}(t) = \hat{s}_{Tr}\left\{\underbrace{\cos(\omega_{Tr}t)}_{\text{载波}} + \underbrace{\frac{\mu}{2}\cos[(\omega_{Tr}+\omega_{Na})t]}_{\text{上边带}} \underset{\text{与AM不同}}{\uparrow} \underbrace{-\frac{\mu}{2}\cos[(\omega_{Tr}-\omega_{Na})t]}_{\text{下边带}}\right\} \tag{3-30}$$

小偏差 FM 的矢量图可由对应的 AM 得到，方法是对下边带矢量旋转 180°，如图 3-30所示。

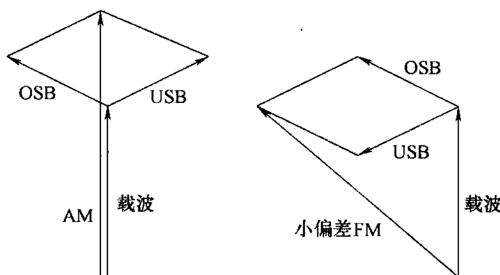


图 3-30 AM（左）和小偏差 FM（右）的矢量图（这里边带矢量的长度画得比较夸张）

小偏差 FM 并没有值得一提的调制增益。它与 AM 或 SSB 相比的优点在于 FM 对幅度非线性的非敏感性。因此可以使用具有很高的有效系数的非线性放大器（放大器类别 C)。这对馈电紧张的地方是很重要的，同样对使用电池驱动的设备（手持对讲机）以及使用太阳能的设备（卫星）也是很重要的。模拟定向无线传输时也使用小偏差 FM，因为在微波频段内，实现一个线性的功率放大器并不是很容易。

小偏差 FM 的带宽效率同双边带 AM 是相同的，像 SSB 那样将带宽减半对 WM 来说是不可能的。

3. 调制器和解调器

FM 信号可以通过振荡器电路中的共鸣电路来产生，这里共鸣电路的共鸣频率会发生变化。这可以通过使用变容二级管（Varicap diode，Varactor diode）代替电容来实现。频率必需随控制电压线性变化，这只能是通过与载波频率有相对较小的偏差来实现。如果要想在线性程度比较好的调制器中得到较大的偏差，可以

1）首先实现一个小偏差 FM，并使用非线性电路对频率偏差进行多倍化处理以得到所希望的值，然后将它与希望的载频进行混频。

2）或将它调制到较高的载频上，然后在希望的频段上进行混频。

FM 调制器也是可以集成的，称作 VCO（Voltage Controlled Oscillator）或 VFC（Voltage to Frequency Converter)。载波通常是脉冲形式的，必须使用 BP 滤波器进行后续处理。

PM 调制器是根据图 3-25 间接地实现的。一种直接的实现方法是让一个 AM 和一个 90°旋转的载波进行叠加（Armstrong 调制器)。它的基本原理是在 3.1.6 节中将要介绍的已调信号的正交描述方法。

在 5.4.2 节中我们将要介绍一种基于 DSP 的纯数字的实现方法。

只要将一个调制器（不仅仅是 FM 调制器）定性地与一个解调器进行反向耦合，如图 3-31 所示，就可以对这个调制器进行线性化。

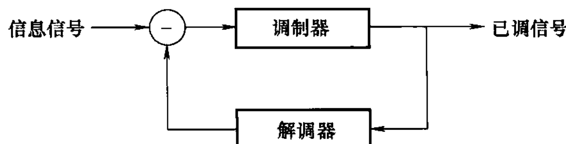


图 3-31 调制器的线性化

如果在发送器和接收器中使用相同的解调器，那么为了实现信号的无失真传输，不需要将它们一次就变成线性的。同样，也可以对图 3-31 中的单元替换，通过一个调制器的帮助来对解调器实现线性化。

PM 解调器由于需要一个参考相位，因此在实际中没被使用。FM 解调器只需要一个参考频率，而这在接收器中很容易产生。回顾一下图 3-25，PM 解调器可以通过一个串联的 FM 解调器和一个积分器来实现。因此我们下面只讨论 FM 解调器。目前存在多种多样的 FM 解调器，更深入的内容可以参见参考文献【Mei92】及【Mäu92】。FM 解调器通常也被称为鉴频器。

FM 解调器通常对如在传输路径中可能遇到的衰减的变化而引起的幅度的变化也很敏感。因此在解调器之前有一个限幅器，将幅度的波动去除掉。

在过于简化的教科书中将限幅器作为 WM 的一个抗干扰能力，解释是干扰只会影响幅度。这是错误的，干扰对零交叉点也有影响。但在较大的偏差，并因此使用较大的传输带宽的情况下，这种错误的影响较小。

在限幅器之后是真正的解调器，众多的方法基本上可以分成 3 类：

1）将 FM 转变成 AM（还包含附加的 FM），然后使用包络解调（截面解调器、比率检测器、频率鉴别器）。该类中不同的方法在早期经常被使用。

2）将 FM 转变成 PDM（脉冲持续时间调制）或是 PFM（脉冲频率调制），然后使用积分器（计数鉴别器、共存解调器）。PDM 和 PFM 将在 3.2 节中进行讨论。

3）带有 PLL（Phase Locked Loop）的相干解调器。这是目前比较喜欢使用的方法，它很容易集成并且可以通过 DSP 的软件来实现。

将 FM 转变成 AM 可以通过例如频率相关的电抗来实现，方法是将 FM 信号放在一个将共鸣频率设置成载频的并行共鸣电路中。这样载频位于共鸣电路的截面上，因此这个解调器也被称为截面解调器。

鉴频器使用微分器来实现由 FM 到 AM 的转换。式（3-19）中 FM 信号幅角的导数为

$$\frac{\mathrm{d}\Psi_{\mathrm{FM}}(t)}{\mathrm{d}t}=\omega_{\mathrm{Tr}}+Ks_{\mathrm{Na}}(t)$$

因此可以得到幅角：

$$\Psi_{\mathrm{FM}}(t)=\int\omega_{\mathrm{Tr}}+Ks_{\mathrm{Na}}(t)\mathrm{d}\tau=\omega_{\mathrm{Tr}}t+K\int s_{\mathrm{Na}}(\tau)\mathrm{d}\tau$$

FM 信号为

$$s_{FM}(t) = \hat{s}_{Tr}\cos\left[\omega_{Tr}t + K\int s_{Na}(\tau)\,d\tau\right]$$

对这个信号进行微分：

$$\frac{ds_{FM}(t)}{dt} = \hat{s}_{Tr}[\omega_{Tr} + Ks_{Na}(t)]\sin[\omega_{Tr}t + k\int s_{Na}(\tau)\,d\tau]$$

COS 函数的内导

这不外乎是一个具有可变频率的 AM 信号，使用简单的包络检测器就可以将 $s_{Na}(t)$ 提取出来。

第二组 FM 解调器首先做从 FM 到 PDM 或 PFM 的转换。以计数鉴别器为例，它以一个很精确的节拍在每个 FM 信号的零交叉点处产生一个宽度和高度都是常数的脉冲。将这个 PFM 信号进行集成，就可以得到一个线性的随频率而变化的电压。

图 3-32 所示是一个锁相环，可以将它与图 2-3 进行比较。它是一个相位调节环，它将 VCO 输出电压的相位与 FM 信号的相位进行比较。与此同时两个信号的瞬时频率也会变成相同。因为带有 VCO 的反方向耦合路径具有 FM 调制器的功能，整个调制器具有相反的功能，即是一个解调器。这相当于图 3-31所示过程的逆过程。相位检测器通常使用平衡混频器（见图 3-5）来实现。这种解调器对幅度不是很敏感，因此不需要一个前置的限幅器。

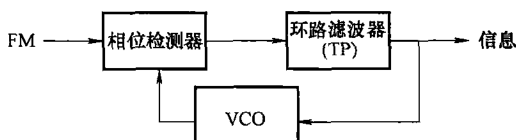


图 3-32 锁相环（PLL）作为 FM 解调器

3.1.5 模拟调制方法的干扰特性

干扰阻抗及调制增益的相关知识对不同调制方法的比较评估及选择正确的调制方法是很重要的。值得一提的是，没有一种方法是理想的，只能说它对于某种给定的应用场合是最优的。

很遗憾干扰特性的计算是很复杂的。为此需要使用数学的方法对干扰进行描述，随机过程理论是必不可少的。这里我们并不进行推导，直接使用它们并对结果进行解释就足够了。计算过程可以参见参考文献【Ste82】、【Mäu92】、【Lük92】、【Kam92】、【Höl86】、【Pro94】。在 2.8 节中已经举例对数字信号的基带传输进行了类似的分析。

图 3-33 给出了评价不同调制方法干扰特性时需要关注的因素。

现在我们假设只有一个噪声，即在传输的信号中叠加一个高斯噪声。这种特殊情况既是很常见的（电路噪声、30MHz 以上无线信道的噪声），同时用数学方法描述起来也相对简单。信道是指调制器的

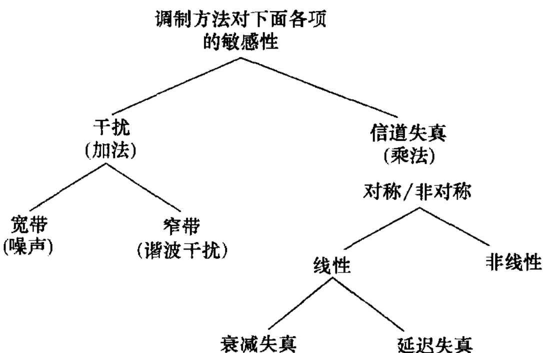


图 3-33 不同调制方法干扰特性的评价指标

出口与解调器的入口之间的这一段，并不仅仅指传输线路那一段。经过这些噪声的影响，在解调器输入端有一个信噪比 SR_K。经过解调器后信息信号有一个信噪比 SR_A，参见图1-43。图3-34所示为不同调制方法时 SR_A 随 SR_K 的变化情况。

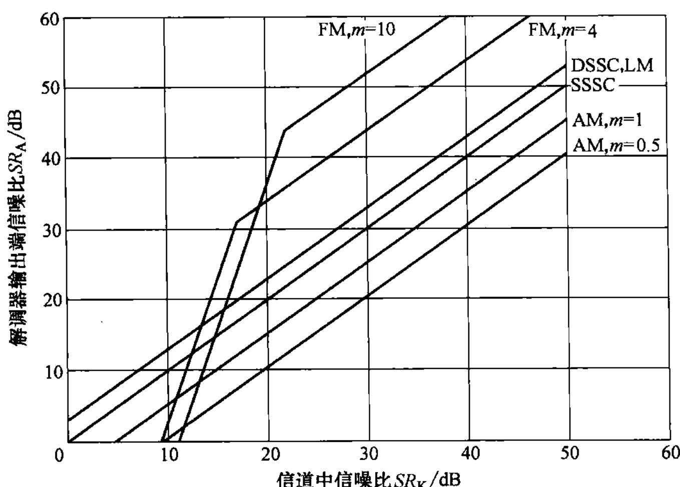


图3-34 不同模拟调制方法的调制增益（FM曲线中陡峭下降部分是按程式化方式描绘的）

SSSC = SSB，它们不产生任何带宽变化，即 $SR_A = SR_K$。因此将SSB的曲线作为一个参考，如果某种调制方法的曲线在SSB的曲线之上，我们就说它有一定的调制增益。

DSSC使用双倍的传输带宽，产生的调制增益是3dB，即系数为2。产生这种结果的原因是，两个边带中的有用部分是相关（等相位）相加的，而干扰部分是不相关（指功率）相加的。或者这样说更清楚：两个边带的信息内容是相同的，但噪声部分并不相同，其中部分可能还会相互抵消。

正常的AM同样使用双倍的带宽，但却没有调制增益。这是因为在信号功率中包含载波，但载波却不包含任何信息。当最大调制深度 $m=1$ 时，AM仍然比SSB低4.8dB。

FM的增益是清晰可见的。对于带宽增大的调制方法（如FM，PCM）来说很典型的是，当 SR_K 的值低于某一固定门限值时，增益曲线下降得很快，可以参见图1-46。增益越大，这个门限值也就越高。通过汽车中使用的UKW接收器我们可以感觉到这个门限值。当开车驶入一个隧道时，开始信号的质量还很好，但驶过几米之后信号突然就中断了。

在接收条件很差（$SR_K < 15$dB）的情况下，也要考虑一下SSB、AM和FM。根据参考表1-8可知，在接收条件极差的情况下，使用FM根本不能进行话音传输，但使用SSB却有可能。

使用包络检测器的AM解调器也有一个门限值。当 SR_K 低于15dB时，AM的情况比图3-34所示的情况还要差。

具有一定的带宽放大系数 β 的调制增益理论上是可以得出来的，如图1-45所示。图3-35把图1-45中的曲线以及图3-34中的曲线SSB（$\beta=1$），$m=1$ 时的AM（$\beta=2$）

及 $\mu=4$ 时的 FM （$\beta=10$） 放在了一起。从这里可以清楚地看到，现有的调制方法与最大可能工作值还相差较远。只有 SSB 可以作为理想情况来描述，但它没有任何增益。FM 的增益随着带宽的增大只是某二次方的增长，但理论上指数的增长是可能的。结论：

对干扰的保护最好使用信道编码，而不是在调制中进行！

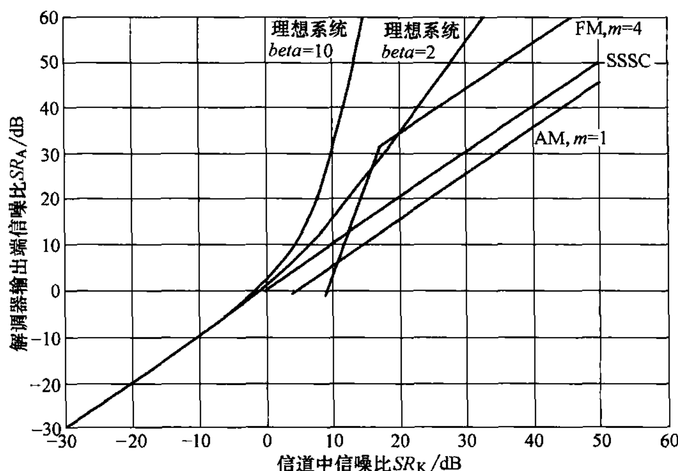


图 3-35 理想的（长曲线）与现实的（短曲线）调制方法的比较

下面我们用窄带噪声干扰信号来代替宽带噪声干扰。这种情况在同时存在两条无线信道，以及使用相邻的频率进行发射等都会出现。此时不希望的次要波会干扰希望得到的接收信号，另外例如在军事应用中对有用信号进行故意的干扰也属于这种情况。

上面提到的 DSSC 相比 SSSC 有 3dB 的增益仅适用于噪声干扰。在窄带干扰的情况下，SSB 更有抵抗力。这是因为在传统 AM 以及 DSSC 情况下的非对称边带对解调后的信号会产生令人很不舒服的像敲击声一样的非线性失真。因此在长时间拥挤的短波波段，已经介绍过的 ECSS 解调器有很大的优点。一个很明显的原因是：如果两个边带受到轻微的并且不同的干扰（如噪声干扰），两个边带的平均值会比单独一个边带的值要好。如果一个边带受到了很大的干扰（窄带干扰），那么没被干扰的边带比平均值更接近于原始信号。

在 FM 的情况下，我们要观察一下所谓的遏止效应（capture effect）：如果干扰大于有用信号，那么会根本收听不到有用信号。这种特性应用在同频同播网络中，如果要想在较大的范围内对手持无线对讲设备进行覆盖，那么就要在山上设置中继站（接收-变频-发送）。中继是双工的，而无线对讲设备是半双工的。因此变频是必不可少的，这样中继站的发送器才不至于将他自已的接收器阻塞了。不外乎是根据 FDD 原理进行方向分离。现在不对网络的所有中继站都分配一对频率，而是让所有中继站都工作在相同的频率对上。无线对讲设备在不同的地方从发射功率最强的一个中继站收听信号。

但这对处于从不同站点收到相同信号强度的边界地带会出现困难。为此将中继站点的发送频率设置 10～15Hz 的偏差。这样在边界地带会有一个带有该偏差的拍频干扰。这样做话音信号虽然会有些切割声，但还是可以理解的。这样的同频同播网络节省频率信道，也能简化设备的操作，因为它一直使用相同的频率。跟其他问题一样，使用它也有一个障碍：同频同播网比较容易受到大面积的干扰。

我们还需看一下信道失真的影响。信道失真可以分成线性/非线性、对称/非对称失真。这里对称是针对载频而言，非对称失真的起源例如可以是没有正确调整的接收滤波器。

先来看一下线性失真。在实际中 WM 对衰减失真是完全不敏感的，但信道的延迟失真会在解调器输出端引起非线性失真。如果对带宽的限制很大，也会因此非线性失真。只要卡森（Carson）带宽能够被满足，那么谐波失真就会小于 1%。

SSB 在解调器之后会引起线性失真（但这只是将 HF 频谱推到基带），但音色并不会改变。

传统的 AM 受对称信道失真的影响会引起线性失真。非对称信道失真会引起乘法解调器的线性失真，而引起包络检测器的非线性失真。选择性衰减（即两个边带衰减不同，短波情况下有可能会出现这样的问题）绝大多数情况下会引起非线性失真。

非线性失真可以很快地被处理：WM 对它是不敏感的，而所有类型的 AM 对它是敏感的，并且会引起非线性失真。特别棘手的情况是交叉调制，即调制时邻居干扰载波被传到有用载波上，发送器好像突然“用错了频谱”。原因跟互调制一样，在于接收器输入端的不同级别输入不希望有的非线性信号。

3.1.6 已调信号的正交描述

谐波形式的载波可以对幅度和/或相位进行调制，参见式（3-2）。值得注意的是所有的调制类型都可以归纳成两种幅度调制后信号的和。

利用式（3-2）：

$$s_m(t) = a(t)\cos[\omega_{Tr}t + \varphi_{Tr}(t)]$$

以及三角公式：

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

于是得到：

$$\begin{aligned} s_m(t) &= a(t)\cos[\varphi_{Tr}(t)]\cos(\omega_{Tr}t) - a(t)\sin[\varphi_{Tr}(t)]\sin(\omega_{Tr}t) \\ &= s_k(t)\cos(\omega_{Tr}t) - s_q(t)\sin(\omega_{Tr}t) \end{aligned}$$

在最后一行得到两个新的信号，即所谓的

同相成分： $$s_k(t) = a(t)\cos[\varphi_{Tr}(t)] \tag{3-31}$$

正交成分： $$s_q(t) = a(t)\sin[\varphi_{Tr}(t)] \tag{3-32}$$

很明显，两种成分都来源于已调信号的包络 $a(t)$ 及零相角 $\varphi(t)$。进行反向推导，有

$$a(t) = \sqrt{s_k^2(t) + s_q^2(t)} \tag{3-33}$$

$$\varphi(t)=\arctan\left[\frac{s_q(t)}{s_k(t)}\right] \tag{3-34}$$

由于反正切函数的解不是唯一的，$\varphi(t)$ 的值在 $k\pi$，$k=0$，±1，±2，…中并不确定。因此对已调信号式（3-2）的另外一种表示方法为

$$s_m(t)=\sqrt{s_k^2(t)+s_q^2(t)}\cos\left[\omega_{Tr}t+\arctan\frac{s_q(t)}{s_k(t)}\right] \tag{3-35}$$

式（3-35）中没有考虑的幅角的不确定性在实际中常常是不重要的，因为通常人们只是对包络感兴趣，或是幅角的范围通过其他的途径已经是已知的。

式（3-35）的表示方法同复杂交流电技术领域对信号的描述方法具有相同的形式。那里实谐波电压信号：

$$u(t)=\hat{U}\cos(\omega t+\varphi)$$

用复数来表示为

$$\underline{u}(t)=\underline{U}e^{j\omega t}$$

这里 $\underline{U}$ 的含义是

复数的幅度： $\underline{U}=\hat{U}e^{j\varphi}$

原始的实信号 $u(t)$ 来源于复信号 $\underline{u}(t)$，这是通过实部构造的方式，即在实轴上的投影来实现：

$$\begin{aligned}u(t)&=\mathrm{Re}[\underline{u}(t)]=\mathrm{Re}(\underline{U}e^{j\omega t})=\mathrm{Re}(\hat{U}e^{j\varphi}e^{j\omega t})\\&=\mathrm{Re}[\hat{U}e^{j(\omega t+\varphi)}]=\hat{U}\mathrm{Re}[\cos(\omega t+\varphi)+j\sin(\omega t+\varphi)]\\&=\hat{U}\cos(\omega t+\varphi)\end{aligned}$$

这与式（3-2）所描述的已调信号完全类似。从

$$s_m(t)=a(t)\cos[\omega_{Tr}t+\varphi_{Tr}(t)] \tag{3-36}$$

中可以得到：

$$\underline{s}_m(t)=\underline{a}(t)e^{j\omega_{Tr}t} \tag{3-37}$$

在式（3-37）中复数的幅度是与时间有关的，因此人们也将它称作

复包络：
$$\begin{aligned}\underline{a}(t)&=a(t)e^{j\omega_{Tr}(t)}=a(t)\cos[\varphi_{Tr}(t)]+ja(t)\sin[\varphi_{Tr}(t)]\\\underline{a}(t)&=s_k(t)+js_q(t)\end{aligned} \tag{3-38}$$

复包络曲线 $\underline{a}(t)$，它的值是实包络曲线 $a(t)$，它的实部是同相成分 $s_k(t)$，虚部是正交成分 $s_q(t)$。

也可以由式（3-37）通过构造实部的方式反过来得到式（3-36）：

$$\underline{s}_m(t)=\underline{a}(t)e^{j\omega_{Tr}t}$$

$$\begin{aligned}s_m(t)&=\mathrm{Re}[\underline{s}_m(t)]=\mathrm{Re}\{[s_k(t)+js_q(t)]\cdot[\cos(\omega_{Tr}t)+j\sin(\omega_{Tr}t)]\}\\&=s_k(t)\cos(\omega_{Tr}t)-s_q(t)\sin(\omega_{Tr}t)\end{aligned}$$

为什么需要这些呢？使用复数方法进行描述的意义包括以下两点：

1）理论上的研究：前面所介绍的带通信道和带通信号可以通过等价的低通信道和低通滤波器来描述。这种重新描述需要引入频率搬移的概念，因为调制来源于傅里叶

变换理论，因此只研究TP信道就可以了，结果可以推广到BP信道。

2）实际中的实现：所有的调制方法在基带中都能实现，然后在希望的带通设备中进行转换。

用数字的方法来实现这些概念是很容易的。在5.2.2节中我们将再次讨论这点。

3.2 脉冲载波的模拟调制

在非线性很强的信道中（例如二元信道，只允许被传输的信号有两个状态，甚至会出现磁滞现象）传输模拟信号，必须对它进行调制。这要用到脉冲形式的载频。但是调制后的信号并不是数字的，因为它的幅度和时间并不同时是离散的。调制后信号的频谱比原始信息信号的频谱大很多，理论上存在很陡的脉冲截面甚至是无限宽的。已调信号的频谱从0Hz开始，它涉及的是一种基带方法。

模拟信息信号的脉冲调制可以分成以下几种方法：（1）PAM（脉冲幅度调制）、（2）PWM，PDM（脉冲-宽度-调制，脉冲持续时间调制）、（3）PPM（脉冲相位调制，脉冲位置调制）、（4）PFM（脉冲频率调制）、（5）RFM（矩形波频率调制）。

信息信号是模拟的，所以它的时间和数值都是连续的。承载的信号是脉冲形式的，所以只是在特定的时间才会出现。因此信息信号在被调制之前必须经过采样处理。结论：

1）载波的脉冲频率必需足够高，它必须要高到能满足采样定理的要求，参见式（1-2）。

2）接收器在解调之后要使用低通滤波器来重新构造出信息信号。

图3-36是不同模拟脉冲调制方法的比较，说明如下：

1）PAM几乎与AM相同，它的唯一优点是：脉冲占空系数（脉冲持续时间与脉冲周期的比值）可以做得很小，因此多个PAM信号可以同时传输（如时间复用）。当然短脉冲需要较大的带宽。

2）PWM、PPM、PFM和RFM的脉冲高度相同，信息包含在时间轴上。因为绝大多数干扰都会影响幅度，而很少会影响较陡的脉冲截面，因此这些信号对传输路径中的干扰具有较好的抵抗力。同时这些信号也很适合在非线性系统中处理。它们可以应用在如模拟光传输（PFM和RFM）及电路交换网络（PWM）中。

3）多数情况下PPM是通过PWM来产生的。它可以用在下降沿或是图3-36所示的PWM的上升沿使用。

4）PFM和RFM的区别是：PFM的脉冲宽度是常量，而RFM的脉冲占空系数是常量。对RFM使用带通滤波器可以过滤出基波，因此它就像一个普通的FM。

5）对于时间复用来说PAM是最适合的，因为每个脉冲边缘的时间点都是已知的，但很遗憾PAM对干扰的抵抗力并不是很好（可以通过PCM进行补救）。

6）PAM主要是中间阶段产物（例如采样之后），很少用来传输。

7）脉冲调制方法是一种基带方法，即调制到任意的频段上是不可能的，因此需要对它进行二次调制（参见3.4节）。

8）图3-36所示的调制方法都不是数字的。PAM情况下信息包含在脉冲的幅度上，

它的幅度值是连续的但时间轴是离散的。所有其他的调制方法信息都是包含在时间轴上，时间是连续的，但脉冲的幅度是离散的。数字信号要求时间和幅度同时是离散的。

因为所有这些调制方法只有在远距离传输才被使用，一般使用得很少，所以我们不详细叙述了。感兴趣的读者可以参见参考文献【Mäu91】和【Höl】。

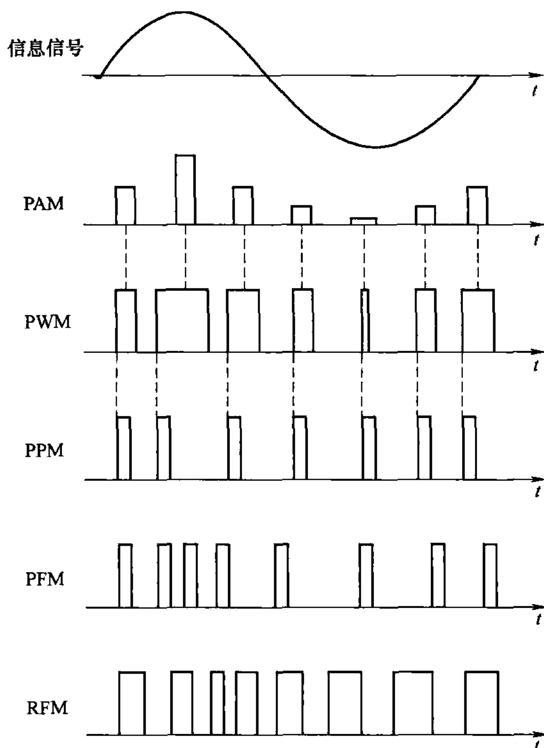


图 3-36　模拟脉冲调制方法

3.3　脉冲载波的数字调制

由于信宿解析能力的限制以及最终涉及到的领域会有所不同，每个模拟信号可以由通过对数值进行离散化，即使用含有相同的信息量的量化信号来代替。

香农已经证明，根据采样定理，带宽受限的模拟信号可以用包含有相同信息量的时间离散的信号来代替。

如果将上面两种情况结合起来，那么就会有时间和数值都是离散的信号，即数字信号。数字信号的优点在 1.1.3 节中已经进行详细的描述。

从应用者的角度来说，下面的因素是互相关联的：

1）抗干扰能力：需要有尽可能少的量化等级。

2）保持原样：需要尽可能多的量化等级。

3）经济适用：数字信号比较容易进行时间复用。

采样（时间离散）信号在基带使用脉冲序列，描述起来很方便。因此 3.2 节中所描述的脉冲调制方法对传输数字信号来说是比较方便的。上面的最后一条显然与幅度离散的 PAM 信号比较相近，因为这种情况下所有的时间边界点都是已知的。但它却不能同时满足前面两种情况。这种困境使用一种编码技术就可以得到解决（参见 1.1.3 节），这种编码技术就是脉冲编码调制（PCM），它不仅对传输起着很大的作用，对信号处理也很重要。PCM 的基本原理在 1939 年由 A. H. Reeves 提出。

3.3.1 脉冲编码调制

1. PCM 的工作原理

对信号处理来说，PCM 无外乎就是持续不断的对信号进行模/数转换。PCM 过程主要分成 3 步，如图 3-37、图 3-38、图 1-7 及图 1-8 所示。

1）采样：即进行所谓的采样—保持—替换。这里注意需要满足采样定理的要求，必要时在采样—保持之前应使用抗失真滤波器（anti-aliasing filter）。该步骤的结果是产生一个 PAM 信号。

2）量化：使用量化器，如级联比较器来实现。由应用来决定量化的粒度。经过量化之后，会有部分信息不可挽回地丢失。如果量化粒度足够小，该步骤发生的只是减少不相关的信息。这一步的结果是产生一个量化的 PAM（QPAM）信号。

3）编码：使用编码器，即一个逻辑电路来实现。绝大多数情况下会编成一个二元码（对偶码，二元补码等）。结果是产生一个 PCM 信号，可以将它认为是二元的 PAM。这样每个采样值会用多个二元脉冲来传输。这些脉冲必须足够短，因此会占用很多带宽。这样做的结果是提高了抗干扰能力，这又涉及到对信息立方体进行重塑的问题，如图 1-35 所示。

多数情况下，采样、量化、编码都是在同一个集成电路中完成的。



图 3-37 脉冲编码调制的 3 个步骤

PCM 这些好处所带来的代价是：

1）带宽需求。

2）量化时信息的丢失。

3）设备的复杂化。但这一点在今天来说并不是很重要，因电子元器件已经变得很便宜。

PCM 脉冲串的基带传输采用第 2 章所述的办法，带通传输将在 3.4 节中进行介绍。对 PCM 脉冲串来说，只要知道它的源就可以了，它的含义并不重要。数字传输系统只是负责传输数字信号。

在 3.3 节的后续部分我们将介绍如何将模拟信号转变成这样的数字信号。关键词是减少量化噪声（非线性量化、压扩）、减少数据传输速率（预测编码）及不必要的传输带宽、误码率的影响及传输增益。

2. 量化噪声

量化即取整的过程，可以看成是叠加误差或噪声信号的过程，如图 1-34 所示。如果采样频率与模拟信号没有确定的对应关系，误差信号的瞬时值是随机的，与噪声信号的特性一致。因此也称之为量化误差。

从图 1-8 及图 3-38 可以看出，误差信号的幅度最大可能是量化间隔的一半。

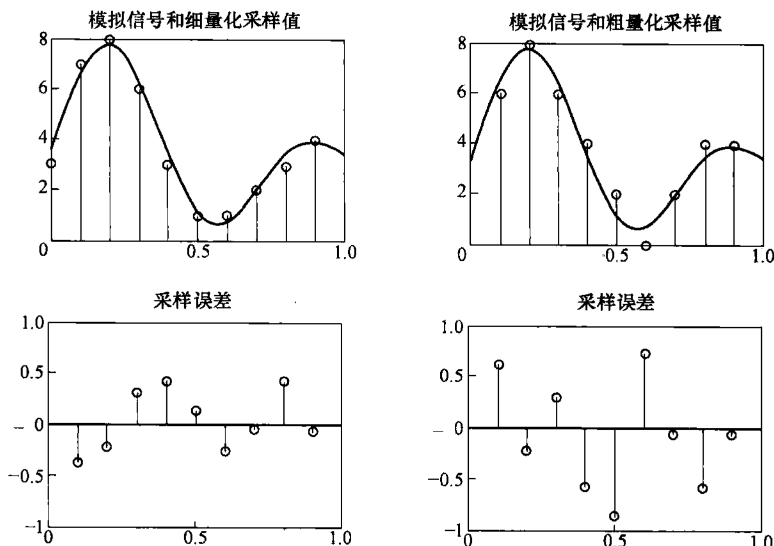


图 3-38 细粒度与粗粒度量化的区别

如果将 A/D 转换器的字宽加大一倍，那么转换器的量化级别数会翻倍，而误差信号的幅度会减半，功率会减小为 1/4。而实际中量化信号的功率不会随量化级别数而发生变化（如果这种说法成立的话，级别越多越好）。单凭经验来说，量化噪声间隔 SR_Q，信号功率 P_S 与量化噪声功率 P_Q 的比值为

$$SR_Q = 10\lg\left(\frac{P_S}{P_Q}\right) \approx 6k \quad (\text{dB}) \tag{3-39}$$

式中 k——A/D 转换器的位数（字宽）。

k 位转换器共可以描述 2^k 个不同的幅度等级，分别是 $0\cdots 2^k \cdot q$（V）或 $\pm 2^{k-1} \cdot q$（V），q 是量化间隔。如果一个模拟信号，如音乐具有 80dB 的动态范围（可通过的最大声与最小声之间的差），那么可以根据下面的方法计算 A/D 转换器的所需的字宽：

$$80\text{dB} \triangleq 10000:1 \Rightarrow 2^k \geqslant 10000$$

$$k > \log_2(10000) = \frac{\lg 10000}{\lg 2} = 13.29$$

$$\Rightarrow k = 14$$

那么需要一个至少是 14 位的转换器。根据经验公式（3-39）可以得出相同的值。

$$k \geqslant \frac{80}{6} = 13.3 \Rightarrow k = 14$$

第三种计算方法是使用信道能力公式。具有 20kHz 带宽及 80dB 动态范围的信号根

据式（1-17）或式（1-18），它的信息含量为 $I=533\text{kbit}$（注意是小写的 b!）。在没有冗余的编码情况下，对这些信息进行描述需要 533kBit（大写的 B!）。每秒钟都要积累 533kBit 的数据，这些数据根据采样定理分布在 40000 个采样值上。每个采样值描述 13.3 个 bit，即需要 14 个 Bit。注意在最后一步的计算中，采样频率不能高于最低要求的 40kHz。由于模拟信号的带宽限制在 20kHz，过多的采样值不包含任何新的信息，只是冗余。

量化噪声功率不能进行统一的计算。如果模拟信号和误差信号是不相关的，即采样频率与信号是无关的，那么误差信号 $e(t)$（e = Error）的幅度分布是均匀的。量化噪声误差的概率密度分布如图 3-39 所示，这里 q 表示量化间隔。整个曲线的积分是 1（图 3-39 中直方波的面积）。

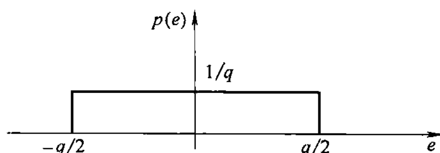


图 3-39 量化误差的概率密度

概率计算的术语中信号 $s(t)$ 的平均值叫做期望值 $E(s)$。它与概率密度是相关的：

$$\text{平均值：}\quad E(s)=\overline{s(t)}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}s(t)\,\mathrm{d}t=\int_{-\infty}^{\infty}sp(s)\,\mathrm{d}s \tag{3-40}$$

信号的有效功率是信号二次方值的均值，它与信号的概率密度是相关的：

$$\text{平均功率：}\quad P_{\mathrm{S}}=\overline{s^2(t)}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}s^2(t)\,\mathrm{d}t=\int_{-\infty}^{\infty}s^2p(s)\,\mathrm{d}s \tag{3-41}$$

式（3-40）和式（3-41）的积分函数中包含信号的瞬时值 s 和它的概率 $p(s)$。上面的公式对所有的静态和遍历的随机过程都是适用的。统计学家称式（3-40）和式（3-41）分别为统计一阶矩和二阶矩。后者（即功率和它的有效值）与平均方差关系密切。

式（3-41）对所有的信号 $s(t)$ 都是通用的，现在将它代入到误差信号 $e(t)$ 中。它的功率与量化噪声的功率是相同的。根据式（3-41）与图 3-39 有：

$$P_{\mathrm{Q}}=\int_{-\infty}^{\infty}e^2p(e)\,\mathrm{d}e=\int_{-q/2}^{q/2}e^2\frac{1}{q}\mathrm{d}e=\frac{e^3}{3q}\bigg|_{-q/2}^{q/2}=\frac{q^2}{12}$$

$$\text{量化噪声的功率：}\quad P_{\mathrm{Q}}=\frac{q^2}{12} \tag{3-42}$$

式（3-42）的前提条件是静态的和遍历的随机过程，以及附合图 3-39 的幅度均匀分布的误差信号。如果一个 1kHz 的正弦信号使用 4kHz 的采样频率进行采样，误差信号本身也是周期性的，它的幅度分布遵循另外一种特性。对实际中常用的情况式（3-42）的结果是很好的。值得注意的是，信息信号和误差信号的实际曲线不一定总会出现特例，这正是使用概率计算的好处。

现在又可以使用式（3-42）来计算量化信号的信噪比。以一个正弦信号被完全调制的情况为例：一个字宽为 k 的转换器共有 2^k 个幅度级别可以用来表示信号幅度范围

内的正、负值。因此正弦信号的峰值最多可以为 $2^{k-1}q$。它的功率为有效值的平方，即

$$P_S = \left(\frac{2^{k-1}q}{\sqrt{2}}\right)^2 = 2^{2k-3}q^2$$

作为干扰信号，误差信号的功率可以根据式（3-42）来计算。将这个功率转换成 dB 就是所要求的量化信噪比：

$$SR_Q = 10\lg(12 \times 2^{2k-3}) = 10\lg(15 \times 2^{2k}) = 10\lg(15) + 20k\lg(2) \quad (\text{dB})$$

$$SR_Q = 1.76 + 6.02k \quad (\text{dB}) \tag{3-43}$$

这样一个字宽为 14 的转换器的量化信噪比为 86dB，字宽是 15 位时它的值是 92dB。根据经验公式（3-39）所得的结果分别是 84dB 和 90dB。字宽每增加一位会得到 6dB 的增益在两种情况下都是成立的，但这里还存在一个常数。这个常数与信号的形状以及 A/D 转换器的调制幅度有关，因此必须不断地重新计算（如果想知道精确值的话）。这个常数与所谓的 Crest 因数，即峰值（标准情况下用调制幅度表示）与有效值（标准情况下用功率表示）的比有关。因此式（3-39）可以进一步细化：

$$SR_Q = 6k + K \quad (\text{dB}) \tag{3-44}$$

正弦信号在完全调制的情况下，$K = 1.76$。对感兴趣的随机信号应该尽量避免有很高的调制幅度，否则有可能会出现很高的瞬时值，引起转换器的过调制。这会引起非常不舒服的非线性失真。调制幅度的经验公式为：

$$\frac{B}{4} = \frac{q \cdot 2^{k-1}}{2^2} = q \cdot 2^{k-3!} = \sqrt{P_S} = \text{信号的有效值} \tag{3-45}$$

这里 B 是 A/D 转换器调制幅度范围的一半，即信号幅度的最大值。

对于数字随机信号的信噪比，由式（3-45）和式（3-42）可得

$$\frac{P_S}{P_Q} = \frac{q^2 \cdot 2^{2(k-3)}}{q^2/12} = \frac{12 \times 2^{2k}}{2^6} = \frac{3}{16} \times 2^{2k}$$

$$SR_Q = 10\lg\left(\frac{P_S}{P_Q}\right) = 20k\lg(2) + 10\lg\left(\frac{3}{16}\right) = 6k - 7.3 \quad (\text{dB}) \tag{3-46}$$

由此可见，丢失的比 1 位字宽还多。

如果一个转换器的调制幅度过小，MSB 被保留，并且可能有其他位没有改变，即出现信息丢失，因此量化信噪比会很差。相反过调制会引起转换器的过饱和，会引起很大的非线性失真。

总结如下：

量化信噪比：$SR_Q = 6k + K$ （dB）
其中 k = A/D 转换器的有效使用字宽
K = 常数，与信号的形状有关
正弦信号：$K = 1.76$
随机信号：$K = -10 \sim -6$

参考表 1-7 给出的信噪比，可以得出表 3-2 所示的 A/D 转换器字宽的参考值。

表 3-2 与应用相关的 A/D 转换器字宽的参考值

信号内容	A/D 转换器字宽	量化信噪比 SR_Q/dB
话音	8	40
音乐	16	90
视频	8～12	40～60

根据表 1-8，当字宽是 1 位时（也就是说，A/D 转换器是一个比较器，用来确定采样时刻瞬时值的符号），话音信号几乎是可以听懂的。而实践也验证了这个令人惊讶的猜测。

3. 非线性量化（压扩）

话音信号的幅度分布如图 3-40 所示[Mau91]。这并不是如图 3-39 所示的均匀分布，而是一个拉普拉斯分布。

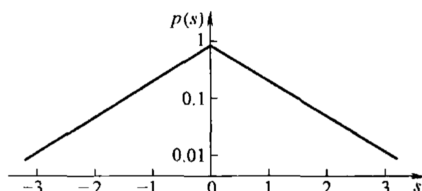


图 3-40 话音信号的概率密度
（归一化的幅度分布，半对数描述）

这里高幅度很少，A/D 转换器并没有得到充分的利用。对模拟信号进行预校正，使幅度分布变成均匀分布可以改善这种情况。理想的分布是如图 3-39 所示的情况。预校正不外乎就是像在 1.3.2 节中所介绍的那样，对信号的瞬时值进行压缩和拉伸。

图 3-41 给出了线性与非线性量化的对比。从量化误差的特性中可以看出，非线性量化的优点是很明显的：对经常会出现的较小的瞬时值来说，误差是很小的；而很少会出现的较大的瞬时值的误差较大。或者换一种方式表达：线性量化给信号增加了一

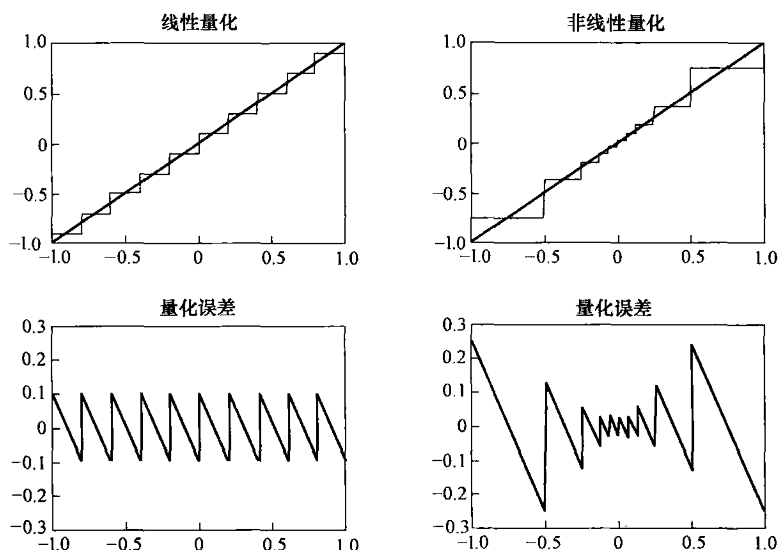


图 3-41 线性量化（左）与非线性量化的比较
上：模拟信号（斜坡）和量化后的信号 下：量化误差

个绝对值是常数的误差，而非线性量化给信号增加了一个相对值（百分比）是常数的误差。

当量化信噪比是常数时，非线性量化的特性曲线即压缩曲线是对数曲线，它穿过零点。ITU-T 建议了两种特性曲线，即欧洲使用的 A 特性曲线及在美国使用的 μ 特性曲线。

A 特性曲线：

$$y=\begin{cases}\operatorname{sgn}(x)\dfrac{A|x|}{1+\ln A} & 0\leqslant |x|\leqslant 1/A\\ \operatorname{sgn}(x)\dfrac{1+\ln(A|x|)}{1+\ln A} & 1/A\leqslant |x|\leqslant 1\end{cases} \tag{3-47}$$

μ 特性曲线：

$$y=\operatorname{sgn}(x)\frac{\ln(1+\mu|x|)}{\ln(1+\mu)}\quad 0\leqslant |x|\leqslant 1 \tag{3-48}$$

A 特性曲线对非常小的信号来说是线性的。A 特性曲线的增益可以达到令人骄傲的 24dB。这里增益意味着提高小输入信号时的信噪比，或提高调制幅度的范围。

在标准的 PCM 传输系统中，参数 $A=87.56$，$\mu=255$。图 3-42 给出的是 A 率、μ 率及 13 折线特性曲线。它们之间的差别小得可以忽略。

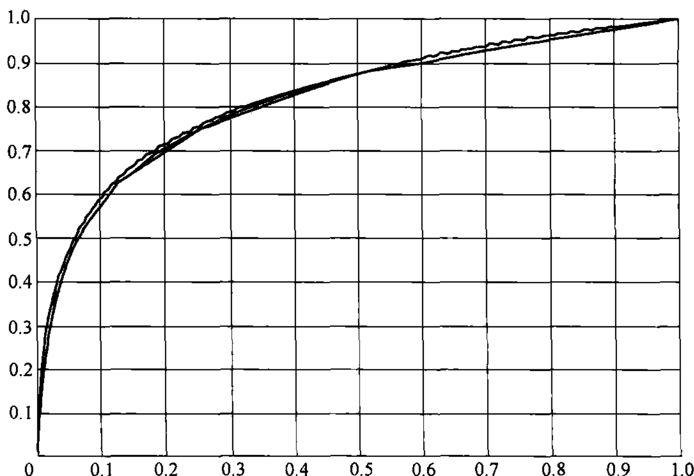


图 3-42 A 率、μ 率及 13 折线特性曲线的比较
（这里给出的只是第一象限的值，第三象限的值是点对称的）

出于实践的要求，A 特性曲线用一个被称为 13 折线的多边线段来表示，图 3-43 给出了它的结构。

特性曲线被分成不同的线段，每条线段的斜率在不断地增加。这些线段被称为 A、B…G。线段 G 通过原点进入到第三象限，并且在第一和第三象限有相同的斜率。因此在第一和第三象限总共有 13 个具有不同斜率的线段，这就是这种特性曲线名称的来源。

每条线段都有 16 个量化等级，线段 G 除外。G 有 32 个量化等级，因此它被分成 G Ⅰ 和 G Ⅱ 两段。在第一象限共有 $6\times16+1\times32=128$ 个量化间隔，整个曲线共有 256 个

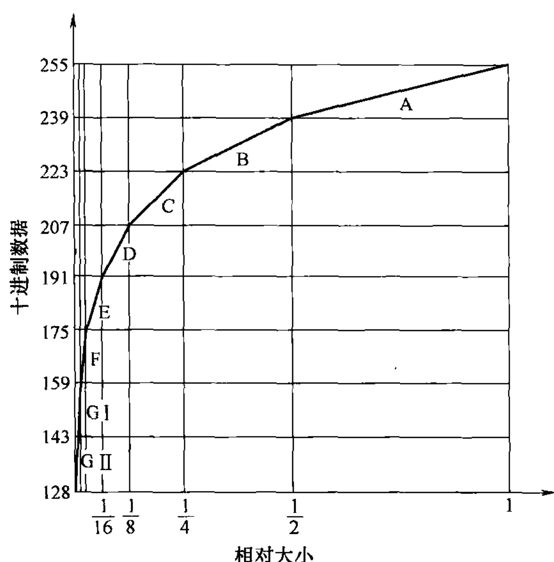


图 3-43 13 折线压缩特性曲线（只是第一象限的值，第三象限的值是点对称的）

量化间隔。因此可以使用 8 位的转换器。

如果将 DC 信号的幅度从满调制幅度到半调制幅度进行变化，线段 A 将被抹掉，这样 A/D 转换器输出值就线性地减少 16 个等级。如果还想将幅度的值减少相同的数量，即出现幅度值 0，那么线性的 A/D 转换器还会将 16 个等级的值抹去。由于对称性，在负半平面也有 32 个等级。可以这样认为，存在一个具有 64 个等级，即字宽为 6 的 A/D 转换器。

现在来重复这个实验，但开始值是调制幅度的一半，然后减少 50%。同样 A/D 转换器也会减少 16 个等级，但这次是减少线段 B。线性外推法的结果还是需要一个字宽为 6 位的 A/D 转换器。

显然对所有的幅度值可能都需要 64 个量化等级，因此量化信噪比与调制幅度无关，是一个常数，并且根据式（3-44），随着信号形状的不同，信噪比的值在 36 ~ 40dB 之间。图 3-44 给出了使用 13 折线量化信噪比的曲线。

现在我们来观察图 3-44 中的加粗线。它是使用正弦信号，13 折线调制时得到的测量曲线。在幅度很小时它是一个线性转换器（线段 G），因此曲线从斜率为 45°开始线性增加。因为根据式（3-44）有一个双倍的幅度（增加 6dB）因此 SR_Q 也增加 6dB。

在调制幅度很大的情况下，曲线 a 在水平方向上波动。在第一个弯处，调制幅度横跨线段 G 和 F。线段 F 的输入信号很弱，当幅度增加时它会调节得更好，SR_Q会增加 6dB。从 A ~ F 总共可以有 6 个波纹，每个波纹从比前一个低 6dB 的跳变处开始，然后以 45°的斜率线性增涨。

信号幅度为 0dB，会出现过调制，即出现非线性失真。

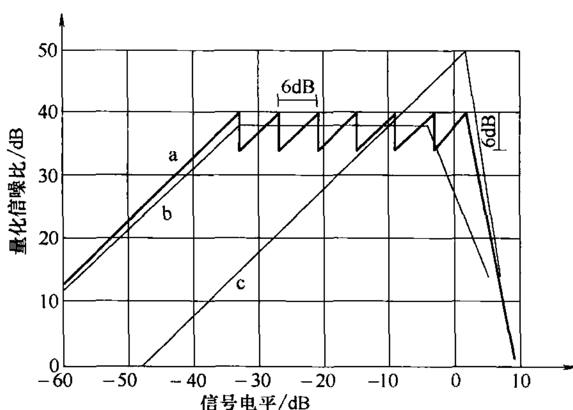


图 3-44 13 折线量化信噪比曲线

a—使用正弦信号控制

b—使用窄带噪声控制

c—线性 8 位转换器，用正弦信号来控制

曲线 b 是当用例如带宽为 300 ~ 550Hz 的窄带噪声信号驱动转换器时产生的。然后在 800 ~ 3350Hz 这个范围对噪声进行测量，计算量化信噪比。这种方法在实践中会用到。

最后折线 c 表示的是字宽为 8 位的线性转换器的量化信噪比，是用正弦波驱动测试得到的。根据式（3-43）在曲线的末尾能达到大约 50dB 的信噪比。不过根据表 1-7 这 50dB 信噪比比话音传输所必需的要好。

另一方面人们也知道，当幅度是 -40dB 时，即 1% 的调制，使用压扩功能，信噪比也能超过 30dB，而采用线性量化该点的信噪比只有 9dB。

从声音上直接进行对比，也可以看出当调制很小时，非线性 A/D 转换器明显的比线性 A/D 转换器好。在大信号的情况下，线性 A/D 转换器的劣势不是很明显。

有多种方法可以实现压扩器：

1）模拟预失真然后线性量化，缺点是需要应用模拟电路原件。

2）A/D 转换器进行非线性量化。这需要特殊的转换器，但这完全是可以实现的。

3）先使用大字宽（至少 12Bit）转换器进行线性量化，然后进行非线性编码转换成 8Bit。实践中这可以采用纯数字的方式实现，如图 3-45 所示。

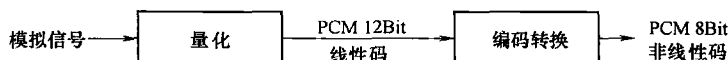


图 3-45 通过编码转换实现压扩（数字压扩）

13 折线的压缩增益是 24dB，这在图 3-44 中曲线 a 和 c 的最大垂直距离处可以看到。线性转换器在输入幅度较小时至少也应该做到这样，因此根据式（3-44）需要 24/6 = 4Bit 额外的字宽，即转换器的字宽应为 12Bit。

压扩器 8Bit 的的数据字是按照下面的方法来分配的：

1）1Bit 给符号位；

2）3Bit 分配给段（图 3-43 中算上 G Ⅰ 和 G Ⅱ 共有 8 段）；

3）4Bit 分配给段内的位置（总共 16 个级别）。

除了通过压扩的办法可以改善量化信噪比之外，还有一个概念可以实现相同的效果，即过采样。它的想法是这样的：量化噪声是个随机函数，它的功率频谱在 0 到采样频率的一半即 $f_A/2$ 之间平均分布。因此噪声功率遵循式（3-42），与采样频率无关。需要关注的只是模拟信号的频率范围 B_a。如果采样频率设置得远高于香农边界频率 $f_A = 2B_a$（这也是过采样这个名字的来源），那么量化噪声功率的一部分就会分布在需要关注的频率范围之外，可以使用数字的办法利用低通滤波器将它过滤掉。这样信噪比就会得到改善，如图 1-67 所示。

这是可以实现的，因为模拟信号的带宽限制在 B_a。附加的采样值不包含任何新的信息，它们是冗余的，也就是说相互是没有关联的。量化时对采样值进行四舍五入，有的值被减少，有的值被增加。由于 TP 滤波器相互的依赖性，四舍五入的误差可以部分相互抵消，这样信噪比得到了改善。

在双倍的过采样（$f_A = 4B_a$）中，噪声功率在 $0 \sim B_a$ 范围内与按最小采样频率采样（$f_A = 2B_a$）相比减少了一半。这样信噪比在传输速率翻倍的情况下改善了 3dB。然而此时量化等级数也翻倍（即转换器的字宽增加 1 位，从 k 变成 $k+1$），这样信噪比在传输速率提高了系数 $(k+1)/k = (1+1/k)$ 的情况下改善了 6dB。因此过采样对提高信噪比来说并不是很经济的办法。然而它还是得到了应用，不过不是应用在传输技术中，而是用在了信号处理上。它的优点是模拟抗失真滤波器和重构滤波器的边缘无需很陡峭。

3.3.2 预测编码

1. 差分 PCM

如果一个信号完全控制 A/D 转换器并且使用香农边界采样频率进行数字化，那么 n 时刻的采样值可能在正调制幅度（Ausstererungsgrenze）附近，而 $n+1$ 时刻的采样值可能在负调制幅度附近。在两个相邻的采样点之间没有任何关联，PCM 信号的各位是不可预测的值。

相反采样值之间相互关联意味着至少部分采样值是可以预测的，它们是冗余的，不再包含最大可能的信息量（1Bit < 1Bit，即 1Bit 的信息内容小于 1bit）。在下面的情况下，PCM 信号前几个 Bit 变化的不如后几个 Bit 频繁，因此可以进行具有一定程度准确性的猜测。

1）信号只包含低频成分，但转换器却是满调制的。补救方法是减小采样频率。

2）部分调制。补救方法是使用基准适配来对调制进行改进。

3）信号的低频成分幅度较大而高频成分的幅度较小。补救措施是使用差分 PCM（DPCM）。

话音和音乐信号的情况跟上面最后一种情况正好相似。频谱分析表明，高音的幅度明显小于低音的幅度。最主要指导原则是采样频率。结果是两个采样值之间的差小于最大可能出现的信号幅度的两倍。

差分 PCM 的想法是，不传输采样值，而传输两个采样值之间的差。将差值用与原始信号相同的解析率来描述，使用较小的字宽，例如用 4 位代替 8 位，就足够了。并且这样数据流也会相应减少，即 DPCM 也能减少冗余。

DPCM 在图像传输中也被广泛地使用，因为除了场景发生变化之外，时间相连的各幅图像的内容都很相似。

声音信号的特点： 高幅度的成分很少 →通过压扩来提高信噪比 高频成分很弱： →通过预测编码来减少冗余

2. 预测编码的原理

预测编码的原理是发送器发送的不是单个的采样值，而是它与预测值之间的差值。这个预测值是已知的，可以通过前一个采样值来计算。然后将预测值与实际值的差通过正常的 PCM 向接收器传输。在接收端需要像在调制端一样增加一个预测器，它根据已经收到的并且已经解码的值计算出相同的预测值。这样两个预测器都有相同的估计偏差，但这两个偏差都用相同的差值信号进行纠正，如图 3-46 所示。

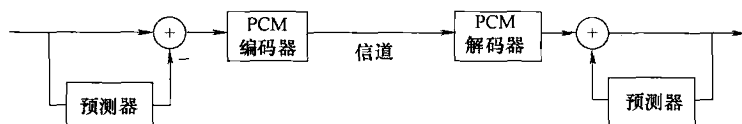


图 3-46 预测编码原理

这样预测的值越准，差值信号所需的字宽就越小。在 DPCM 中简单地将上一个采样值作为预测值，因此预测器实质就是一个延迟器。更精确的办法是不只使用上一次的采样值，而是根据前两次的采样值推算预测值。这也可以推广成通过计算梯度（用割线近似代替阶梯近似）来进行预测。根据前 n 个采样值，使用 $n-1$ 阶多项式计算出一个近似值。但这对节省几个字宽来说过于复杂。

带有预测编码的传输系统受传输误差的影响会引起误差扩散。通常有两类补救措施。

1）每隔一段时间传输整个采样值，用它来代替与预测值的差值。当然需要对该值进行标识，因为这会打破原有的节奏。

2）使用信道编码，通过增加冗余的办法提高抗干扰能力。当然这会导致有用数据传输速率的下降。

图 3-46 所示的原理图有一个值得改进的不足：发送端中的预测器的输入信号是模拟的，而非量化的信号。相反在接收端中，预测器的输入信号只有量化后的信号，因此它们的估计值会有一定的偏差。图 3-47 所示的原理图是对这种情况的改进，这里对发送端进行了修改而接收端保持不变。这样发送端的输出完全可以得到像图 3-46 中接收端那样的编码信号。这样两个预测器就有相同的预测值。这种反耦合原理一般情况

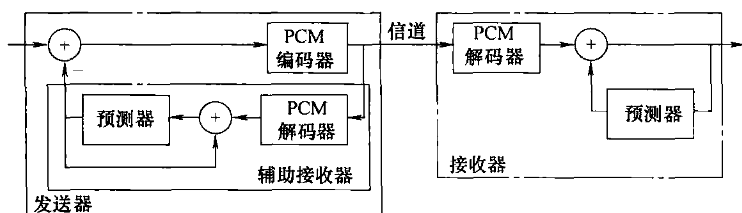


图 3-47　对发送器进行改进后的预测编码

下可用于改善调制器，例如用于调制器的线性化，比较图 3-31。

3. 增量调制（DM）、自适应 DM 和 Sigma-DM

增大 DPCM 的采样频率，超过采样定理所要求的最小值，那么就会减小与预测值的差值，进而减小 DPCM 信号所必需的字宽。在极端的情况下一个 Bit 就够了。DM 就是字宽为 1 的 DPCM。它只传送模拟信号是比预测值大还是比预测值小这个信息。

DM 与 DPCM 相比的最大优点就是它的字宽只是 1Bit，这样就不需要字同步。DPCM 与 PCM 的最大区别是采样频率不再是根据客观的要求即采样定理来确定，而是根据主观的要求，即信号的质量来规定。增量调制时采样频率与量化等级的数目（幅度分辨率）有关。

DM 与 DPCM 类似，预测器实质上就是一个延迟器。相反编码器是一个触发器，它同时也是采样保持单元。图 3-48 所示为对图 3-47 的发送器进行的修正。

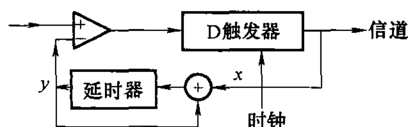


图 3-48　对图 3-47 所示的 DPCM 发送器进行修正后用于 DM

下面我们计算图 3-48 中的信号 $x[n]$ 和 $y[n]$：

$$y[n]=x[n-1]+y[n-1]\Rightarrow Y(z)=X(z)\cdot z^{-1}+Y(z)\cdot z^{-1}$$

对 $Y(Z)/X(Z)$ 进行求解，可得由加法器和延迟器组成的串联反馈电路的传输函数：

$$Y(z)\cdot(1-z^{-1})=X(z)\cdot z^{-1}\Rightarrow\frac{Y(z)}{X(z)}=\frac{z^{-1}}{1-z^{-1}}$$

根据信号处理技术我们知道最后面的表达式是一个延迟、脉冲不变的积分器。据此可以得出如图 3-49 所示的 DM 调制器和解调器的原理图。接收器输出端的低通滤波器可以去除脉冲频率中的遗留部分。

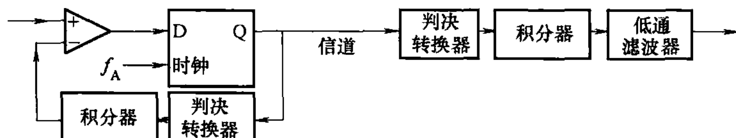


图 3-49　DM 调制器和解调器原理图

图 3-49 中的积分器可以用数字（与输出端的 D/A 转换器结合起来）或模拟的方法来实现。两个积分器都需要使用双极性脉冲来控制，脉冲高度就是阶跃高度，因此

需要两个判决转换器。相反信道中的脉冲序列可以是单极性的，并且高度可以是任意的。这些脉冲只需通知积分器到来的是正脉冲还是负脉冲。

另外一种简单的硬件实现方法如图 3-50 所示，它使用上/下计数器，后接一个 D/A 转换器。

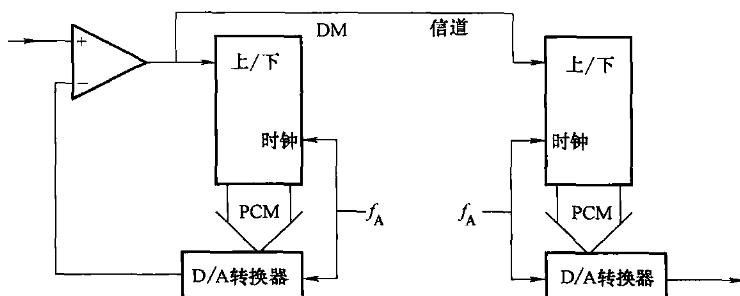


图 3-50 使用上/下计数器的轨迹 D/A 转换器作为数字增量调制、解调器

由于 DM 比较容易实现，它对许多方面，特别是军事系统中的话音传输有很重要的意义。

当然增量调制也有它的缺点。图 3-51 描述的是一个有问题的模拟信号及其在 DM 传输系统输出端重构时情况。

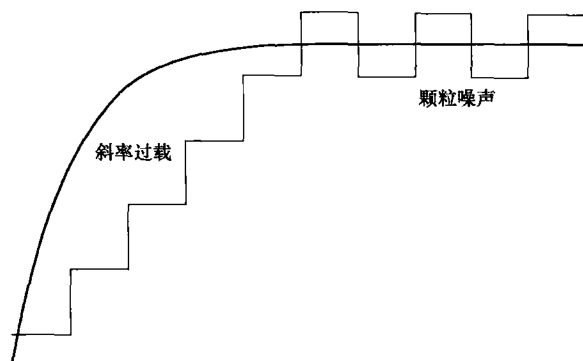


图 3-51 增量调制的信号特性

这里有两点值得注意：

1）持续不断的输入信号只是用交替变化的符号来描述。这种由 DM 系统产生的信号波动称为颗粒噪声（granular noise)，它能引起话音停顿干扰。3 个差值（正、零、负）的传输是不可靠的，因为它需要三元码或一个字同步系统。将量化级差高度减小会缓和这种问题，但又会增强下面的问题。

2）如果输入信号的变化很快，当量化级差高度和/或采样频率很小时，编码后的信号与原始信号可能不一致。这种现象称为斜率过载。补救措施包括：提高采样频率（不好，因数据传输速率会提高)、增大级差高度（不好，会提高颗粒噪声)、自适应增量调制（好)、Sigma 增量调制（好)。

自适应增量调制（ADM）

ADM是指如果差值的符号多次保持不变的话，就增大量化级差高度。如果符号发生变化了，就再减小级差高度，如图3-52所示。使用这个协定（当然要进行准确的定义），调制器和解调器可以独立地设置各自的预测器。这可以通过改变图3-49中的判决转换器来实现，如同在DM中传输二元信号那样。

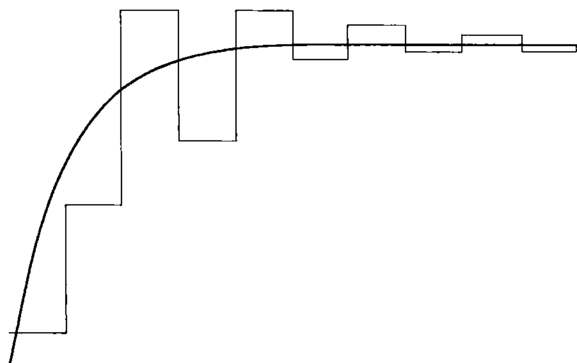


图3-52 ADM的信号特性曲线

ADM也被称为“高信息增量调制（High Information Delta Modulation，HIDM）”、“连续变化斜率增量调制（Continuously Variable Slope Delta Modulation，CVSD）”或“数字控制增量调制（Digitally Controlled Delta Modulation，DCDM）”。

像正常DM那样，根据瞬时级差高度的不同，ADM也会引起较大的传输误差。跟所有的预测调制方法一样，也需考虑到误差扩散的情况。对信息信号特性的了解会帮助改善误差扩散的状况。不仅对ADM来说是这样的，DM和DPCM的情况同样如此：众所周知，声音信号是不包含直流分量的，但传输误差可能会引起直流分量，所以可以对已解码信号的均值进行规定和调节。

Sigma增量调制

通过对模拟信号进行预失真，可以避免斜率过载，因而不会引起较大的信号改变。线性失真用积分器来实现比较有优势。在解码器中可以使用微分器来进行补偿。

Sigma增量调制器不外乎就是一个积分器加上一个增量调制器，Sigma增量解调器就是一个增量解调器加上一个微分器。增量解调器本身由一个积分器和一个低通滤波器组成，参见图3-49。级联的积分器—低通滤波器—微分器可以使用一个独立的低通滤波器来实现，如图3-53所示。与图3-49进行比较可以看出，这只是简单地将积分器从解调器的输入端移动到调制器的输入端。图3-54给出的是一个简化的发送器，它将图3-53的两个积分器结合在一起。

如果信息信号为零，那么就会传送相同的正、负脉冲。信息信号值为正，会引起过多的正脉冲。脉冲序列的均值相当于要寻找的信号值，这个均值使用低通滤波器从脉冲序列中提取出来。

DPCM、DM、ADM及SigmaDM中数据的减少实际上是减少冗余，它的基础是充分

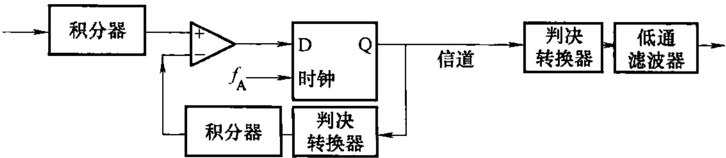


图 3-53 Sigma 增量调制器和解调器

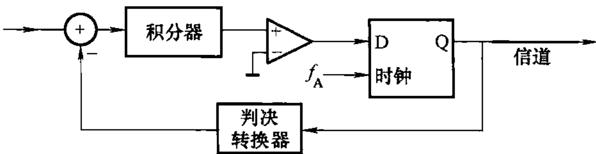


图 3-54 图 3-53 中简化的发送器

利用事先对信息信号的了解（与4.1.1节进行比较）。如果事先不了解信息信号（如话音信号高幅度的成分很少，高频成分很弱），那么数据传输速率与PCM相比并不会减少。尽管如此，人们还是很喜欢用SigmaDM，原因是它的解码器特别简单。不仅如此，在高质量的信号处理中，Sigma增量调制器也得到了越来越多的应用。在A/D转换器中，信号的采样频率远高于采样定理所要求的频率，可以使用1Bit编码来描述。与PCM不同，真正的A/D转换器仅仅使用一个比较器。由于有了过采样（oversampling），抗失真滤波器的边缘不需要非常陡。此外，量化噪声的很大一部分位于感兴趣的信号频带之外（比较3.3.1节第3部分中结尾一段）。使用额外的负反馈，可以实现量化噪声频谱的主要部分集中在高频处，即超出信息信号最高频率处（噪声整形），这可以进一步改善信噪比。使用这种方法，A/D转换器的动态性可达120dB，相当于PCM情况下的20位字宽。而可以使用PCM来进行处理。编码转换使用数字Sigma增量解调器来完成，即3-53中的低通滤波器使用数字的办法实现，然后减少采样速率。

相反在D/A转换中使用插值来提高采样速率，转换成Sigma增量码。使用模拟低通滤波器可以恢复出模拟信号。这个滤波器也有重构滤波器的作用，但并不需要因为采样频率过高而具有很陡的边缘，因此很容易实现。一些CD设备通常使用这种方法，但并没有令人误解地称之为1Bit转换器。

最后我们在表3-3和表3-4中比较一下各种数字基带调制方法。这里所给出的传输速率是近似值。

表 3-3 话音传输所用数字基带调制方法的比较

调制方法	字节	采样频率	复杂性	话音的传输速率
PCM	大	低	中	64kBit/s
DPCM	中	低	大	32kBit/s
DM	小	高	中	32kBit/s
ADM	小	中	中	16kBit/s
SigmaDM	小	中	小	16kBit/s

表 3-4　传输前增加信号带宽的数字 BB 调制方法的适合性

评判标准	适合性
频率转换	不可行(要求其他的调制方法,如 FSK 等)
信道调节	所有的信息都可调节
复用	时分复用是可行的,频分复用只有配合使用其他方法(如 FSK 等)才可行
复杂性	与模拟调制相比复杂性较大 使用积分器使复杂性的区别变得较小 信道没有被很好利用的,变得较昂贵
对信道的要求	较小的干扰,较大的带宽

在 ISDN 中，话音传输使用的是 64kBit/s 的压扩 PCM，而不是节约性的方法。原因是 ISDN 不仅传输话音业务，还有数据业务。因此减小 ISDN 信道的数据传输速率是没有意义的。

3.3.3　PCM 的传输带宽

传输 PCM 信号时，每个采样值用 k 个 Bit 来表示。对采样频率 f_A，下面的香农定理成立：

$$\frac{1}{T_A} = f_A > 2B_a$$

式中　B_a——模拟信号的带宽。

当每个采样值都用 k 个 Bit 来编码时，每个 Bit 的持续时间 T_{Bit} 在极限情况下的值为

$$T_{Bit} = \frac{T_A}{k} = \frac{1}{kf_A} = \frac{1}{2kB_a}$$

根据式（2-2），二元码传输时最小传输带宽为

$$B_ü = \frac{1}{2T_{Bit}} = kB_a$$

在实践中滤波器的边缘会做到很陡峭（奈奎斯特截面），因此按照下面的方法计算

PCM 传输带宽：　$B_ü = ckB_a$；c 在1.6～2之间　(3-49)

PCM 信号的带宽 $B_ü$ 随着模拟信号的带宽 B_a 的增长而快速增长。因为传输时间没有改变，所以一定表现在抗干扰性上有所提高。这将在 3.3.5 节中进行推导。

计算实例：如果带宽为 3.4kHz 的话音信号以 8kHz 的频率进行采样，并且用 8Bit 进行编码，那么数据传输速率结果是众所熟知的 64kBit/s，采用二元码传输时奈奎斯特带宽为 32kHz。使用滚降系数为 $r = 0.5$ 的脉冲形式滤波器，传输带宽为 48kHz。即带宽的放大系数为 14！当设置 $k = 8$，$c = 1.75$ 时，使用式（3-49）会有相同的结果。

我们使用式（2-5）来计算相同的内容，会产生一个下的差异：

$$B_ü = B_N(1 + r) = \frac{R}{2}(1 + r) = \frac{2kB_a}{2}(1 + r) = kB_a(1 + r)$$

括号内的内容相当于式（3-49）中的 c。我们仍然设 $r=0.5$，那么带宽的放大系数是 12 而不是 14。此时相当于式（3-49）中的 $c=1.5$，而在式（3-49）中我们推荐 c 在 1.6～2 之间。这些附加的带宽是抗失真滤波器的陡峭边缘所要求的。在上面的计算中，出于相同的考虑设置 $f_A=8\text{kHz}$，尽管模拟信号的带宽只是 3.4kHz。

3.3.4 PCM 误码的影响

PCM 信号使用第 2 章所介绍的方法在基带进行传输，此时会产生误码，它的概率在第 2.8 节中已经推导过了。

错误解码的数据产生的影响与应用有关。PCM 情况下使用 LSB 或 MSB 进行码字转换是有区别的，增量调制在这点上比较有优势，对它们并不敏感。

误码对 PCM 话音传输的主观影响可以通过实验得到，结果见表 3-5。

表 3-5 误码对 PCM 话音的影响

BER	主观影响
10^{-6}	感觉不到任何影响
10^{-5}	单个的“喀哒”声，在低音域刚刚能感觉到
10^{-4}	经常会出现“喀哒”声，在低音域能感觉到一点干扰
10^{-3}	不断会出现“喀哒”声，所用话音域都能感觉到干扰
10^{-2}	很强的“劈啪”声，很明显听懂的内容在减少
0.5×10^{-1}	话音很难才能听懂

3.3.5 PCM 的调制增益

经常出现的“喀哒”声跟量化噪声一样很明显。假设每个编码字最多错了一个比特，总共只有几个比特错误，那么根据参考文献【Höl86】计算出的信噪比与量化级别的数量无关，即

$$\left.\frac{P_S}{P_N}\right|_{dB}=SR_A=10\lg\left(\frac{1}{4p_{\text{Fehler}}}\right) \tag{3-50}$$

它是随误码概率而变化的用 dB 表示的信噪比，是输出端即解调器或 PCM 解码器之后的值。而式（2-24）描述的是误码概率与用 dB 表示的信道的信噪比（缩写为 SR_K）之间的关系，即是解调器之前的值。将式（3-50）与式（2-24）结合起来有

$$SR_A=10\lg\left[\frac{1}{2erfc\left(\frac{1}{\sqrt{2}}10^{0.05SR_K}\right)}\right] \tag{3-51}$$

现在我们可以观察一下 PCM 的带宽投入是怎样转换成抗干扰能力的。我们将推导输出端的信噪比与带宽放大的函数关系。这里信噪比不是用 dB 来表示的。

根据式（3-42）和式（3-45）可以得出量化噪声为

$$\frac{P_S}{P_Q}=\frac{q^2\cdot 2^{2(k-3)}}{q^2/12}=2^{2k}A;\ 0<A\leqslant 1 \tag{3-52}$$

这里系数 A 包含模拟信号的调制幅度及幅度的分布。如果满幅调制时 $A<1$，那么出现小幅度的概率比较大。k 可以根据式（3-49）来计算，将它代入到式（3-52）中有

$$k=\frac{1}{c}\cdot\frac{B_{ü}}{B_{a}}=\frac{1}{c}\cdot\beta_{\mathrm{PCM}}\Rightarrow\frac{P_{S}}{P_{Q}}=A2^{\left(\frac{2}{c}\cdot\beta_{\mathrm{PCM}}\right)} \tag{3-53}$$

这里 β_{PCM} 是 PCM 的带宽放大倍数，A 和 c 都是常数。量化信噪比随传输带宽的线性增长呈指数递增。在 1.1.9 节中已经介绍过，这是带宽与抗干扰能力之间相互转换的最好可能。

PCM 的调制增益很明显。信道的信噪比要求大到能够安全地检测到脉冲即可。根据表 3-5，误码率为 10^{-6} 就足够了，而根据图 2-23 这相当于信道的信噪比 SR_{K} 达到的 14dB。经过解码后，剩下的只是量化噪声，采用线性 8Bit 编码时值为 48dB。因此最大的调制增益为 48dB - 14dB = 34dB。另一方面，对带宽的额外需求当然要求很大。没有调制增益时，$SR_{A}=SR_{K}$，表 1-8 中给出的 14dB 是差得不可接受的。在相同信噪比情况下，PCM 信道本身的已经是很好的。

式（3-51）将 SR_{A} 与 SR_{K} 联系起来，图 3-55 以图形的方式给出了它们之间的关系。

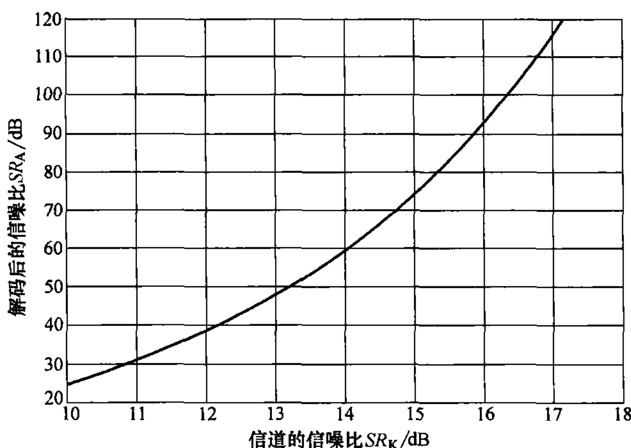


图 3-55 根据式（3-51），SR_{A} 与 SR_{K} 的函数关系

图 3-55 中的 SR_{A} 在误码率为 0 时也不能超过量化信噪比 SR_{Q}，这点是不对的。图 3-56 是对图 3-55 的修正，它考虑了式（3-39）中不同字宽的情况，也与 FM 情况下调制指数 $\mu=5$ 时的调制增益进行了对比。

PCM 解调器输出端的噪声由两部分组成：

1）由误码引起的噪声，是由信道受到干扰后引起的。

2）量化噪声，它总是存在的。

当 SR_{K} 低于 12dB 时，输出端的总噪声主要是由信道的误码率决定的；而当 SR_{K} 高于 16dB 时主要是由 PCM 的字宽决定。这中间是个过渡阶段，SR_{K} 每提高 1dB 就会使 SR_{A} 提高 10dB。当 SR_{K} 大于 16dB 时，SR_{A} 就不再提高，这可以用图 1-46 中提到的发送

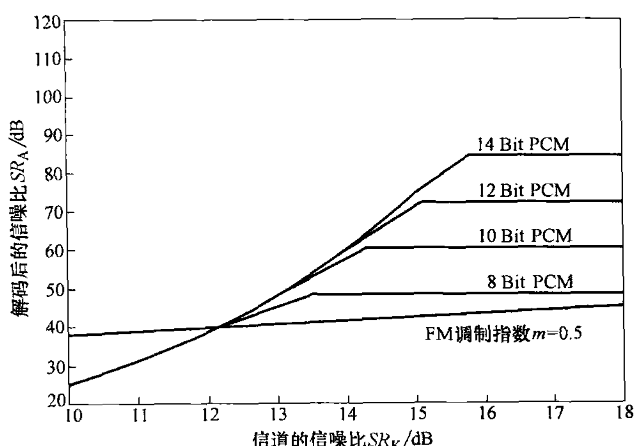


图 3-56 PCM 与 FM 的调制增益

功率的调节来解释。

所有的数字调制方法都很典型地具有波动效应，因此需要规定信道信噪比的均值，在这个均值之上，解调后的信号的质量就很好，如图 1-46 所示。这也说明了，当脉冲能够被识别出来时，它所包含的信息就能够被完全理解。图 2-23 也说明了，信道的信噪比有一点点改善，误码率就会有很大的提高。另外很典型的一点是，当信道的状况很差时（SR_K低于 12dB），FM 好于 PCM。如果将图 3-56 再继续向左扩展，那么 FM 也会有一个弯点，此时 SSB 又会好于 FM，比较图 3-34。

现在我们来看一下 PCM 和 FM 对带宽的需求。对 FM，有式（3-26）：

$$B_{\text{uFM}} = 2(\mu + 1)B_a$$

图 3-56 给出的 FM 的例子中 FM 带宽的放大系数为 12。PCM 的带宽根据式（3-49）为

$$B_{\text{uPFM}} = 1.6kB_a$$

当转换器的字宽 $k=8$ 时，带宽放大系数是 12.8，几乎与 FM 的情况相同。8 位 PCM 曲线与 FM 曲线差别的最大值是 5.5dB。这相当于功率相差 3.5 倍。如果要计算与带宽放大无关的通用的调制增益，可以看到，当字宽为 8Bit 或是更高，也就是说在实际中比较感兴趣的情况时，PCM 明显地好于 FM。

然而 FM 是一种带通传输方法，PCM 却是一种基带传输方法。两种方法不能简单地进行替换。如果将 PCM 信号用 QPSK（quadrature phase shift keying，比较 3.4.4 节）在基带进行传输，所推导出的公式依然成立。

从图 3-56 可以看出，当带宽投入相同时，模拟调制方法比 PCM 的调制增益差。实际这点我们在 3.1.5 节中已经提出了。

现代（数字）信息系统只是为了将频率搬移到需要的频段才使用调制，它使用信道编码来实现干扰保护。

通常人们使用编码增益来代替调制增益，参见 4.3.5 节。

3.4 谐波载波的数字调制

3.4.1 引言

在第2章中我们讨论了数字信号（包括PCM）在基带中传输的情况。图2-11给出了相应的系统组成模块图。图3-57是图2-11的一个简化版本，它省略了发送器和接收器的内部结构。

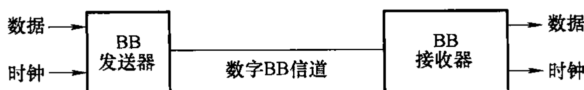


图3-57　基带数据传输

在带通信道中进行传输还需要一个附加的调制，载波采用谐波，如图3-58所示。使用这些调制信号可以很容易地实现频分复用，而使用数字基带信号可以很容易地实现时分复用。载波频率可以很低（例如电话调制解调器）或者很高（如微波频段的无线广播）。

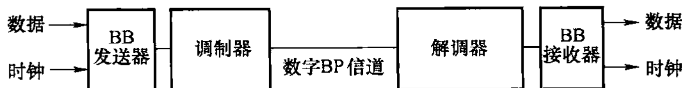


图3-58　BP信道的数据传输

本节要讨论的主题就是图3-58中调制器和解调器两个模块。其他模块在第2章中已经讨论过了，可以直接拿过来使用。关键词像脉冲整形、时钟再生、均衡、线路编码等跟当前的含义相同。但对最后两个关键词需要进行如下的补充说明：

1）BB接收器中的均衡器对信道进行均衡。这里信道既包括传输介质也包括调制器和解调器中频率响应。均衡器可以通过两级来实现，即在解调器之前进行带通均衡，在解调器之后进行基带均衡。

2）BB发送器的线路编码的作用不外乎是将DC分量去掉，因为它会对调制产生影响。因此不使用如HDB-3这样的伪三元码，而是使用扰频器（见图2-6和图2-7下）。

现在我们来看看图3-58中的调制器和解调器。原则上，它正如图3-1所述，是谐波载波的正常调制。因此也需要考虑AM、FM和PM。这里唯一的区别在于信息信号是数字的，即时间和数值都是离散的。相应地载波调制后的参数也变成是离散的。在AM的情况下，将二元数字信号作为信息信号，那么调制后信号的幅度也只有两个值。因此人们不再称它为调制，而是叫它键控，并且区分如下不同的方法：ASK（Amplitude Shift Keying，幅移键控）、FSK（Frequency Shift Keying，频移键控）、PSK（Phase Shift Keying，相移键控）。

为了节省带宽，像在基带传输中那样，不传输直方波，而是传输滤波后的脉冲。脉冲整形在图3-58中的“BB发送器”中进行，这里不再重述。严格说起来，ASK信

号的幅度是连续的，因此有时也使用 ASM（Amplitude Shift Modulation）这样的表达方法。

传输需要的带宽跟在基带传输中相同，都与传输速率（Schrittgeschwindigkeit）有关。如果要使用较小的带宽达到较高的传输速率，那么必须以降低抗干扰能力为代价降低波特率，参见2.3节。调制后的数字信号有多个值，二进制调制也变成多进制调制。调制后数值的个数用数字来表示，如4ASK，8FSK等。在BB传输中通常使用二元或伪三元信号，而在BP传输中经常使用多进制调制。因为随着无线传输在不断发展，HF频谱资源变得很紧张。

在实际中，ASK目前只有在光传输中还在使用。数据传输速率要求较低的情况下由于FSK比较简单因而使用FSK。中等数据传输速率情况下，考虑到带宽的有效性通常选择多进制PSK。而高数据传输速率及多进制调制要求的情况下，会将ASK和PSK结合起来使用，这称之为QAM（Quadrature AM）。

将模拟调制及数字基带传输再复习一下对下面的学习是很有帮助的，因此读者最好将图1-48及其解释概括一下。

选择一种调制方法实际上是寻找一种最优的折中。根据1.1.8节，在每次通信中都有带宽、能量及资金（系统的复杂性）的投入。当然不能同时将所有的需求都降到最低。例如在陆地定向无线传输中，带宽的有效性是最关键的评价指标。而对手机和卫星来说，因为能量与它们的重量有关，因而如何节省能量消耗是最关键的。而对大多数的系统而言，系统的价格是至关重要的。

由于数字技术的高速发展，目前的设备可以做得很复杂、功能很多，但很可靠、价格很低。例如GSM手机（见6.5.2节）和数字电视（见5.1.2节）技术，因此人们几乎可以很轻松地说，降低实现的价格不是什么好办法。另一种表达方式是没有被充分利用的信道比做得很好的发送器和接收器的价格要高许多。使用数字技术会比使用模拟技术更容易实现带宽和能量的有效性。此外，复杂、多功能的数字系统比相同的模拟系统要便宜许多，如图1-47所示。

3.4.2 ASK

图3-59给出了两个二元ASK的例子。当 $m=1$ 时，已调信号的幅度要么为振幅，要么为0，因此人们也将它称为通断键控（On-Off-Keying，OOK）。

ASK信号可以看成是谐波载波和被调制的直方波信号的乘积。它的频谱是两个组成部分频谱的卷积（faltung），结果是以载波频率为中心的 $\sin(x)/x$ 特性曲线。当出现010101串时，$\sin(x)/x$ 曲线两个零值之间的差最大，它的值是 $2/T_{Bit}$，即码元频率的两倍。与基带传输相同，在这个差值的一半处能够正确识别出每个码元。由于是AM，带宽是基带传输的两倍，式（3-15）与式（2-3）在这里也适用。

$$B_{ASK_{min}} = S = \text{波特率} \tag{3-54}$$

由于存在脉冲整形，实际带宽相对式（3-54）要放大1.4～1.5倍，参见式（2-5）。

在实践中ASK只使用二元码。它主要应用在光传输中，发光二极管或激光二极管

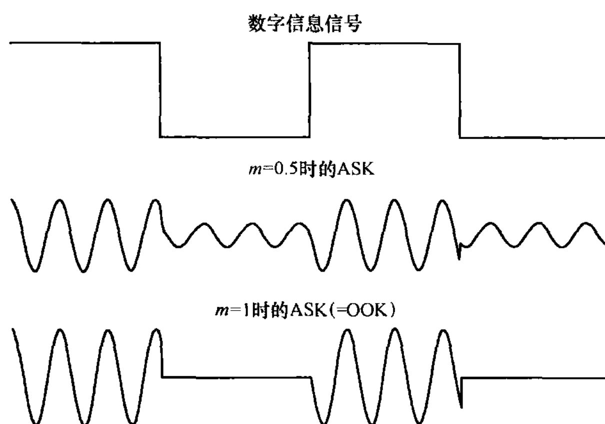


图 3-59　不同调制深度时的 ASK 信号

通过开关来控制光的亮度。由于不同二极管的作用不同，当使用发光二极管时选择 OOK，而当使用激光二极管时是使用 $m \approx 0.9$ 的 ASK。二元 ASK 没有考虑带宽的有效性，因为在光波导带宽是巨大的，这个问题根本不重要。

通断键控在今天也只是被业余无线电爱好者在发送摩尔斯码中使用。这里数据传输速率很低，因此带宽也不是问题。目前的调制大多数情况下都在 800 ~ 1000Hz 范围上，它通过对音频发生器进行采样来实现。然后将每个信号用 SSB 发送器在希望的传输频率上传输。解调器使用 SSB 接收器，这里乘法解调器（见图 3-19）中的振荡器根据接收信号也设置成 800 ~ 1000Hz 范围。如果在发送端发出摩尔斯码，会出现 HF 振荡，解调器就会给出可以听见的差频。

SSB 对数据传输是不适合的，因为基带信号含有低频成分。数字残边带调制（VSB）可发挥作用，因此美国数字电视传输中建议使用该方法。而 SSB 通常用来将第一次调制后的信号搬到另外一个频段上传输。

3.4.3　FSK

在二元 FSK 的情况下，如图 3-60 所示，二进制信息信号将谐波振荡器的频率在 f_1 和 f_2 之间进行切换。

我们定义：

载波频率：
$$f_{\mathrm{Tr}} = \frac{f_1 + f_2}{2} \tag{3-55}$$

频率间隔：
$$\Delta f_{\mathrm{Tr}} = \frac{|f_1 - f_2|}{2} \tag{3-56}$$

根据式（3-22），调制指数为：
$$\mu = \frac{\Delta\omega_{\mathrm{Tr}}}{\omega_{\mathrm{Na}}} = \frac{\Delta f_{\mathrm{Tr}}}{f_{\mathrm{Na}}} = \frac{\Delta f_{\mathrm{Tr}}}{B_{\mathrm{N}}} \tag{3-57}$$

这里 B_{N} 是式（2-3）所给出的奈奎斯特带宽，μ 规定了 FSK 的传输带宽及干扰间

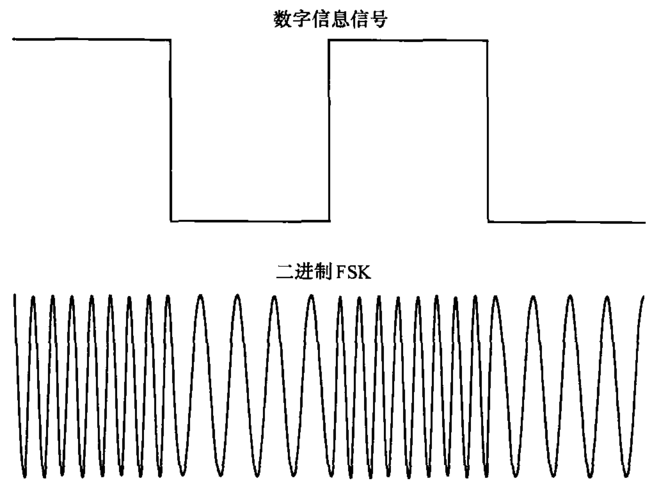


图 3-60 二进制 FSK

隔。一个比较折中的值为

$$\mu = \frac{2}{\pi} \tag{3-58}$$

FSK 信号的带宽不能根据式（3-28）来计算，因为数字信息信号不是谐波。因此人们使用 EXOR 将 2FSK 信号转变成两个 OOK 信号。因此它的频谱是在频率 f_1 和 f_2 处的两个 $\sin(x)/x$ 包络。这在无线数据传输时很有优势，因为与 OOK 不同，在接收器中始终会有信号。因此使用这种方法可以调节无线信道中随时间衰减的信号。

FSK 可以实现多个值，但大多数情况下都使用 2FSK。只有在少数的情况下才使用 4FSK 和 8FSK 系统。

作为一个调制器需要一个压控振荡器（Voltage Controlled Oscillator，VCO）。在两个独立并且固定的振荡器之间进行切换虽然是很明了的，但在实际中由于存在相位跳变而引起的带宽增加而不能被使用。解调器使用在 3.1.4 节中描述的在输出端带有采样器的方法。经常使用的方法还包括所谓的频率转换器，它使用两个带通滤波器将频率为 f_1 和 f_2 的信号提取出来，然后比较它们的幅度，如图 3-61 所示。

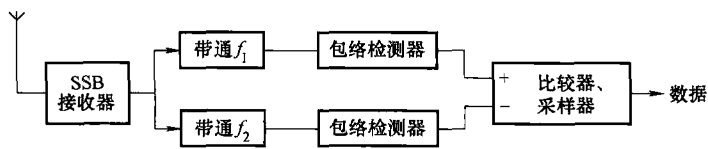


图 3-61 简单的不相干 AFSK 接收转换器

通常情况下，调制是在声频范围内完成的，因此也称之为 AFSK（Audio Frequency Shift Keying）。调制和解调使用简单的运算放大电路或采用 DSP 的方法实现。使用 SSB 可以将 AFSK 信号搬移到实际的传输频率上。同样也可使用 FM 发送器，如常用的手持对讲机，但它在空中使用的是 FM 而不是 FSK。

频率和相位的剧烈跳变会毫无必要地增加已调信号的带宽。因此 FSK 不是离散地而是连续地使用整形后的脉冲实现的。经过对该领域的大量研究，人们得出了不同的节省带宽的 FSK 变种，例如：CPFSK（Continous Phase Frequency Shift Keying，相位连续的频移键控）、MSK（Minimum Shift Keying，最小频移键控）、FFSK（Fast Frequency Shift Keying，快速频移键控）、TFM（Tamed Frequency Modulation，平缓频率调制）、GMSK（Gaussian Minimum Shift Keying，高斯最小频移键控，使用高斯频率响应滤波器进行脉冲整形，主要应用在 GSM 蜂窝移动电话中）。

3.4.4 PSK、DPSK、QPSK、OQPSK

PSK 是数据传输中常用的调制方法。与 ASK 和 FSK 不同，它需要相干解调，正是由于这个原因，在低速数据传输中不值得使用 PSK。二进制 PSK，即 2PSK 是基础，如图 3-62 所示，它会使载波信号的相位发生 180°的跳变。因此也会使用 BPSK（Binary Phase Shift Keying）这个名称。

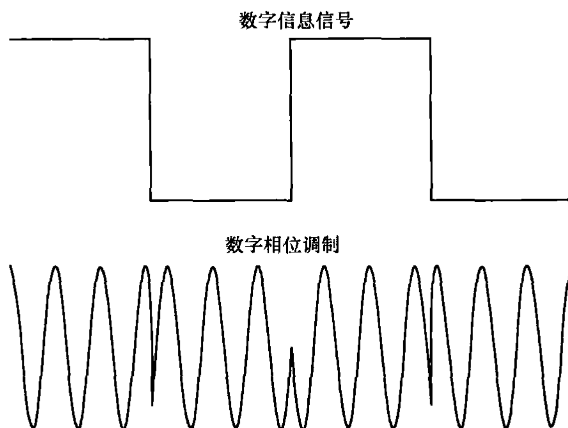


图 3-62 二进制 PSK（2PSK、BPSK）

2PSK 起源于 2ASK，将 2ASK 的调制深度 m 设置得很大。因此载波消失，只剩下两个边带，同混频一样。$m>1$ 时的 AM 是连续 AM 与二进制 PM 的结合，参见 3.1.3 节。如果信息信号是二进制的，那么已调信号的幅度是常数而使用 PM，参见 3.10 节。事实上 180°的相位跳变可以想象成是通过载波与值为 ±1 的脉冲串进行乘法运行而得到的。当与 −1 相乘时，载波就会翻转。实际中 2PSK 通过比较熟悉的平衡调制器来实现，参见图 3-2、图 3-5 和图 3-63。

2PSK 的频谱与 ASK 一样也包含 $\sin(x)/x$ 包络和载频。包络零点之间的距离为 $2/T_{Bit}$。使用一半的带宽（奈奎斯特带宽）进行传输即是可行的，但实际中常使用带宽 $B_{ü}=(1.4\sim1.5)/T_{Bit}$。与 ASK 不同，在 2PSK 频谱的 f_{Tr} 处没有 Dirac 冲击。

解调时需要首先恢复载波，这采用图 3-8 所示的二次方运算后接一个分频器即可。

经常使用的一种二次方运算器是所谓的 Costas-Loop，其他的方法还包括再调制和

带决策反馈的 PLL。进一步的内容可参考相关文献。

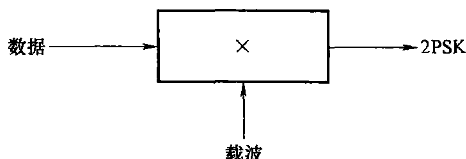


图 3-63 使用混频器来实现的二进制 PSK

PSK 接收器的主要问题是相位同步。图 3-8 所给出的方法由于二次方运算的二义性，可能会产生值为 π 的相位不确定性，当然必须要对它进行纠正。如果接收器从错误的相位开始，那么所有的数据就都被颠倒了。正确的相位信息可以在例如连接建立时使用熟知的同步字来传输，必要时将已解调的数据流进行逻辑的翻转。更好的一种情况是使用差分编码，这样在连接中断之后会实现自动同步。

差分编码时载波和 PSK 信号之间的相位搬移不是由信息（0 或 1 对应于 0 或 π）来选择的，而是根据前一个比特来确定的。对接收器来说测量相位的变化比测量相位值本身简单得多。测量相位还需要一个统一的参考时间点。通常会根据下面的方法分配相位：

正常编码（PSK）：1　相位跟载波比推移 π
　　　　　　　　　0　跟载波相位保持相同

差分编码（DPSK）：1　跟之前的相位比跳变 π
　　　　　　　　　　0　相位不发生跳变

当然相反的分配方案也是可行的。

下面的例子可以更清楚地看到 DPSK 的优点，我们从正常编码开始：

数据：1 0 0 1 0 1 1 1 0 0 1

相位：π 0 0 π 0 π π π 0 0 π

如果接收器恢复的载波相位与调制器中的载波相位相同，那么接收器在第一个比特时就会产生一个相位差，并且将它翻译成逻辑 1。

如果恢复的载波相位相差 π，那么接收器就不会产生相位差，因此检测出的是逻辑的 0。所有整个数据串为

接收器载波正确：1 0 0 1 0 1 1 1 0 0 1

接收器载波相位相差 π：0 1 1 0 1 0 0 0 1 1 0

现在我们来看一下差分编码。第一个比特我们假设发生 π 的差错，后面我们将会看到这个假设一点都不重要。因此传输信号的相位如下：

数据：1 0 0 1 0 1 1 1 0 0 1

相位：π π π 0 0 π 0 π π π 0

接收器需要了解前一个比特的相位。这里我们假设前一个比特固定地为 0，并且接收器第一个有用比特的相位差是固定的，将它解释为 1，相位标记为 π。这样如果下一个比特是 0，就不会发生相位跳变等。整个序列跟希望的是一致的：

开始相位为 0：1 0 0 1 0 1 1 1 0 0 1

现在假设接收器从一个错误的相位 π 开始。那么在第一比特它没检测出相位差别，因此给出 0，相位标记为 π。在第二个比特处仍没有相位跳变，因此仍给出结果 0，标

记相位 π。如此继续下去，整个序列结果为

开始相位为 π：0 0 0 1 0 1 1 1 0 0 1

→从这里开始值是正确的

不管在开始出现哪种问题时，解码的结果都是正确的。因此它适合由同步丢失而引起的短时间中断，差分编码不仅可以实现自同步，还可以实现自愈。

当然差分编码也有缺点：它可能会引起误差翻倍，所有的信道编码都要考虑这个问题。例如使用奇偶位来保护数据然后用 DPSK 来传输是没有意义的。因此使用差分编码的 BER 将比使用正常编码的 BER 大一点，而信噪比相同。

发送器中的差分编码简单地使用异或门与一个 1Bit 延迟器就可实现，如图 3-64 左所示。输出端序列直接送给 PSK 调制器即可。

图 3-64 右侧的电路将接收器中的差分编码重新转换成正常的编码。

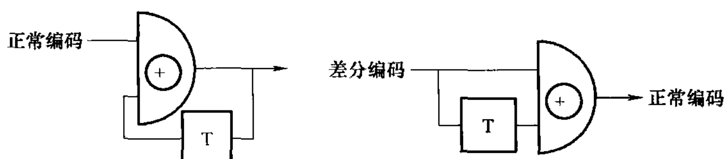


图 3-64 发送器中的差分编码（左）与接收器中的解码（右）

对描述数字调制方法很有帮助的一个工具是星座图，也被称为调制图。这不外乎就是以角速度 $-\omega_{Tr}$ 旋转的已调信号的矢量图，但只是将采样时刻矢量的终点画出来。图 3-65 给出的就是 2ASK 和 2PSK 的星座图。它的两个轴分别称为 I（相位）和 Q（正交），用 Q 轴可表示 3.1.6 节中介绍的正交描述。实际上它就是采样的等价基带信号，参见 5.2.2 节。人们也经常称它为信号空间。

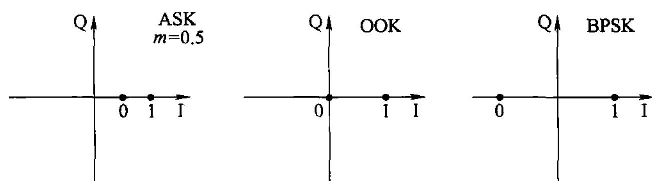


图 3-65 星座图：左：$m=0.5$ 的 2ASK，中：$m=1$ 的 2ASK（即 OOK），右：2PSK（即 BPSK）

从原点到各点的距离是包络的瞬时大小，因此也能表示信号的瞬时功率。

使用 PSK 也能实现多进制调制。4PSK 情况下，有 4 个不同的相位跳变，即 0、π/2、π 及 3π/2。因此它有如图 3-66 所示的星座图。

4PSK 可以看成两个使用正交载波的 2PSK 信号。它实际就是 3.1.3 节中描述的当 $m\to\infty$ 时的正交 AM（QAM）。因此 4PSK 也被称为 QPSK（Quadrature Phase Shift Keying）。一个固定的相位代表两个比特，图 3-66 中给出的是所谓的灰度编码时比特的安排情况。这种安排是很巧妙的，因为解调器最多只能有 π/2 的误差，因此只能引起 1 个比特的误差。

QPSK 和 2PSK 的带宽相同，但在相同的波特率的情况下有双倍的数据传输速率，付出的代价是 QPSK 有较小的抗干扰能力。这从星座图中可以很清楚地看到，QPSK 时每个点的“私有空间”变得较小。图 3-65 和 3-66 给出的是无干扰传输时的星座图。如果受加性噪声的干扰，星座图中的点会跑到错误的位置上。只要没跨过一定的区域，这些点就可以被正确的解码。在接收器中测量到的星座图可以用来评价 BP 的传输质量，相当于 BB 传输中的眼图，如图 2-15 所示。

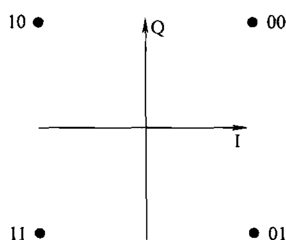


图 3-66 4PSK（QPSK）的星座图

图 3-67 左侧给出的是受到传输干扰时的 QPSK 星座图，右侧是没有传输干扰，但载波再生时有相位抖动的星座图。这也是一种极限的情况。

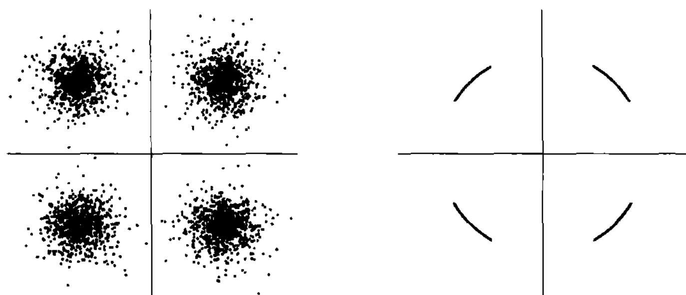


图 3-67 加性噪声（左）及载波再生相位抖动（右）情况下 QPSK 的星座图

星座图中从原点到各点的距离可以表示信号的功率。在相同的最大信号功率情况下，允许进入的调制点越多，每个点的私有空间就越小。相反可以通过增大发送功率来提高私有空间和干扰范围。这在式（1-17）中已经提到过了。

在 QPSK 情况下也可以使用差分编码，同样也可以使用平方运算或 Costas-Loop 来再生载波。

在实践中经常会用 2.6 节中所描述的办法在调制前对脉冲进行整形，这样 QPSK 信号的带宽就不至于那么大。这里值得对 QPSK 进行改进：当相位跳变为 π 时包络会出现快速变化，如图 3-68 所示。但由于带宽的限制，相位的跳变不像图 3-62 那样显著。

使用数学表示式可以很清晰地看出这点。根据式（3-37）和式（3-38），QPSK 信号可以表示为

$$s_{\mathrm{QPSK}}(t) = [s_{\mathrm{I}}(t) + \mathrm{j}s_{\mathrm{Q}}(t)]\mathrm{e}^{\mathrm{j}\omega_{\mathrm{Tr}}t} \tag{3-59}$$

式（3-59）中方括号内的复杂信息信号由 I 信道和 Q 信道两个数据信号组成，对它们进行过滤如图 3-68 所示。同时改变两个信号的状态，那么它们同时穿过零点，即式（3-59）中括号内的部分消失。

当 I 信道和 Q 信道的两个信号不同时改变时，180°的跳变就会避免。也就是说，Q

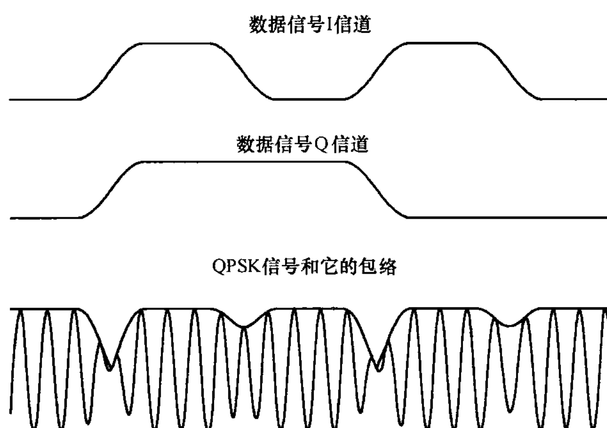


图 3-68 不同相位跳变情况下 QPSK 信号包络的中断情况

信道调制的周期必须是 I 信道的一半。从 00 变到 11 会使 QPSK 经过原点。通过改变时间偏差，这个跳变可由两步来完成：从 00 到 01，再由 01 到 11（或是从 00 到 10，再从 10 到 11），这样每一步的跳变相位变化只有 90°，这种受欢迎的方法被称为 OQPSK（Offset Quadrature Phase Shift Keying），并且它对信道的非线性也不敏感。在该信道中还有一个与调制器相连的功率放大器，这个放大器可以是非线性的，并且有效系数可以做得很高。正因为如此，在卫星技术中喜欢用 OQPSK。图 3-69 给出的是与图 3-68 相同的内容，不过这是针对 OQPSK 的。

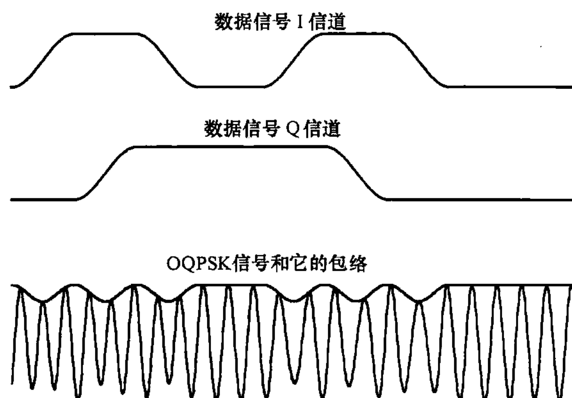


图 3-69 与图 3-68 相同，但用 OQPSK 代替 QPSK

OQPSK 情况下的波特率显然比 QPSK 情况下翻倍了。但我们还记得，QPSK 可以看成是由两个正交的 2PSK 信号组成的，它们的傅里叶变换是线性相加的，并且时间的推移并不会引起幅度谱的变化。因此很明显，带宽并不会因为调制的推移而受到影响。

图 3-70 是 8PSK 时的调制图，所有的点都在一个圆上。

16PSK 也是可以想象的，但此时每个点的私有空间太小，允许的干扰也很小，解

调后的误码率很大。在实际中并不使用16PSK。但为了节省带宽，这样的多进制调制还是很有必要的，因此人们使用QAM来代替PSK。

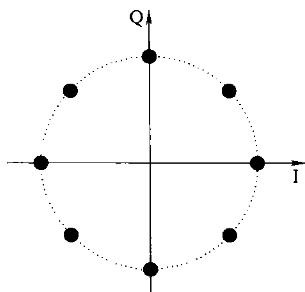


图3-70 8PSK情况下的调制图

3.4.5 QAM

QAM是多进制数字调制方法的统称，在实践中QAM的所有方法都是由PSK和ASK结合起来实现的。

图3-71所示的16QAM星座图中的每个点同具有相同信号功率的16PSK相比都有较大的私有空间。该图所示的位图是这样安排的：前两个比特表示象限，它与图3-66的选择是相同的，即遵循灰度编码。后两个比特表示在某一个象限内部的位置，在第一象限中同灰度编码的安排也是相同的。其他象限内旋转90°后也是相同的。这样安排后，所有的点无论是与水平的还是与垂直的邻居点之间只相差一个比特。对角邻居之间相差两个比特，但由于它们之间距离相差很远，相互混淆的危险很小。

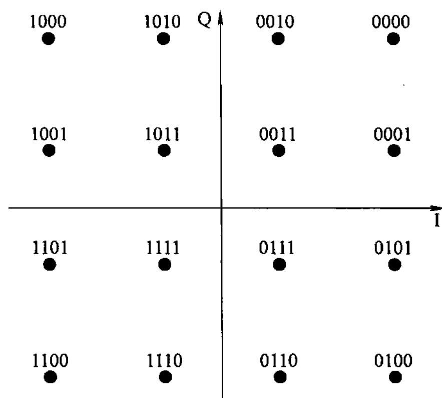


图3-71 16QAM的星座图

图3-72所示为生成QAM的原理图。同QPSK情况类似，也使用I信道和Q信道。如果每个信道都有两个可能的状态，结果就是QPSK，因此被称它为4QAM。如果每个信道有4个状态，那么结果就是16QAM。

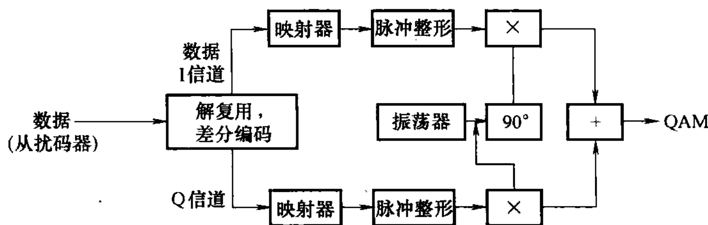


图3-72 QAM调制

图3-72是直接从图3-21得来的。其中两个互不相关的信息信号使用模拟QAM进行传输。这里一个数字化的数据流要以高传输速率进行传输，因此首先必需将它解复用成I信道和Q信道。这实际就是图1-38所介绍的向下复用（Down-Multiplexing）。

解复用之后就是映射器，它根据一组比特来产生脉冲。在4QAM（即QPSK）情况下，每个映射器根据1个比特产生一个正或负脉冲。16QAM情况下，每个映射器根据2个比特产生一个脉冲，幅度可为-3、-1、1、3。

图3-72的映射器之后，为限制带宽而对脉冲进行过滤，接下来便是混频器和将两

个信道进行相加的加法器。

32QAM 情况下，每个映射器产生有 6 个值的双极性脉冲，图 3-73 是相应的星座图。I 信道和 Q 信道各自的 6 种状态共可产生 36 种可能性，它不是 2 的幂。因此人们将星座图中最外 4 个角的点去掉，因为这 4 个点的瞬时功率最大。

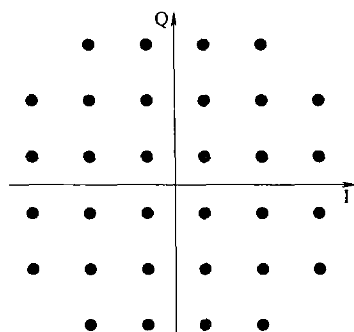


图 3-73 32-QAM 的星座图

实际中甚至可以实现 64QAM 和 256QAM。但它们对信道非失真的要求很高。在接收器中几乎不能省略带有自适应功能的均衡器。

为了使从比特组到调制点的分配比较简单，人们尽量将调制后值的个数设置成 2 的幂。除此之外，I 信道和 Q 信道的状态数应该相同，这样两个信道才可以用相同的方法实现，调制点之间的距离在各个方向上才会相等。

QAM 解调器在从接收信号中提取相位的绝对值时也会遇到困难，因此这里也会使用差分编码。

与 PSK 相比，QPSK 情况下星座图中的每个点与原点的距离是不同的，也就是说，信号的功率根据位图的不同会发生波动。这要求在调制后针对每种情况都进行线性放大，因此在 QAM 中并不使用在 OQPSK 中使用的偏差调制（Offset-Modulation）。

图 3-72 中的所有电路目前都可以使用数字的方法来实现。但这只有当载波频率较低时才是可行的，因此使用 SSB 发送器将调制后的 QAM 信号搬移到希望的频率上进行传输。

一种与 QAM 很相似，有时甚至没有任何区别的方法是 APSK（Amplitude Phase Shift Keying）。此时要用复数指针的模和相位代替它的实部和虚部进行调制。此时星座图中所有点都分布在具有相同中心的圆上。

图 3-74 给出的是所谓的多解析度（Multi-Resolution）-QAM（MR-QAM），它是为图 1-46 所述的数字传输中明显的波动效应而设计的。在 MR-QAM 中，一个象限内的点离得比较近。这样在与正常 QAM 相等的平均信号功率情况下，在每个象限内各点发生混淆的概率会上升，即每 4 个点构成的一个比特组中的后两个比特易出现误码。但前面两个比特发生错误的概率会降低，它们反而被更好地保护起来。与正常的 QAM 相比，随着信噪比的下降，会按顺序出现下面的一系列情况：

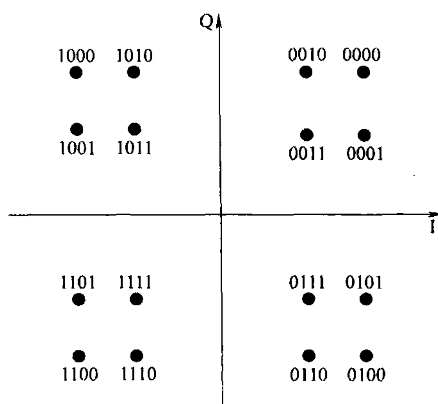


图 3-74 MR-QAM 的星座图

1）MR-QAM 连接质量下降，因为在每个象限内会发生误码。

2）正常的 QAM 连接中断。

3）最后才是 RM-QAM 连接丢失。

MR-QAM 在数字电视（DVB）中得到了应用。调制器如图 3-72 所示，映射器提供相应的修改后的脉冲高度。

误码率采用常用的减少信噪比的办法得到改善。提高发送功率可放大星座图，这样各点之间的距离就会加大。较小的噪声功率会减少每个点周围的云层。

提高 QAM 的码元数目，在传输速率相同的情况下能减少波特率并因此减少带宽，但却以牺牲调制增益为代价；如果相反保持波特率恒定，那么能够传输附加的比特。因此可以考虑用 32QAM 代替 16QAM，使用相同的带宽来传输一个数据流，此时额外的比特可以用作信道编码，即给数据增加一些冗余。多进制调制损失一些调制增益，但信道编码能带来一些已经提到过的编码增益。编码增益的好处要比调制增益带来的损失大，这样整个系统就能接近香农极限。这样做的代价是增加了编码、调制、解调、解码及整个信道适配的复杂性。

在解决干扰保护而不是复用或频率搬移的问题时，调制和信道编码应该结合起来考虑，但是因为两者的理论相差很远，所以到目前为止，大多数研究都是寻找单个情况下的最优解而不是整体的最优解。从整体的角度上看，需要使用新的概念。

根据式（2-6）和式（3-54），在给定传输带宽 $B_{ü}$ 的情况下，数据传输速率 R 为

$$R = \log_2(M)\frac{B_{ü}}{1+r} \tag{3-60}$$

这样用 Bit/s/Hz 为单位表示的频谱效率为

$$\frac{R}{B_{ü}} = \frac{\log_2(M)}{1+r} \tag{3-61}$$

在新的数字电视标准中卷积因子 r 的值为 0.15。这比式（2-6）中推荐的值要小，因此滤波器的边缘要做得比较陡，滤波器会比较复杂。此时调制方法的带宽效率很高，所以将调制器和解调器做得很复杂，同时又省掉脉冲整形滤波器是没什么意义的。64QAM 的频谱效率是 5.2Bit/s/Hz。

图 3-75 所示为误码率与信道信噪比的关系，以归一化的方式给出，即 E_{Bit}/N_0。图 2-24 采用的是相同的描述。误码率的计算方法可以参见参考文献【Kam92】、【Pro94】及【Skl88】。

图 3-75 给出的曲线是最优接收情况下的理想值。实际情况下曲线会稍微向右偏移，主要是因为差分编码会受双倍误码的影响。

数值计算实例：在室温的情况下用 QPSK 传输 2MBit/s 的数据流，滤波器卷积因子 $r=0.35$。当误码率是 10^{-8} 时，接收器输入信号的功率应为多少？

根据图 3-75，这要求 12dB 的 E_{Bit}/N_0。由式（2.26）可得：

$$\frac{E_{Bit}}{N_0} = \frac{P_S/R}{P_N/B_{ü}} \Rightarrow \frac{P_S}{P_N} = \frac{E_{Bit}}{N_0} \cdot \frac{R}{B_{ü}}$$

QPSK 可以使用式（3-61），因此：

$$\frac{R}{B_{ü}} = \frac{2}{(1+r)} = \frac{2}{1.35} = 1.48 \mathrel{\hat{=}} 1.7\text{dB}$$

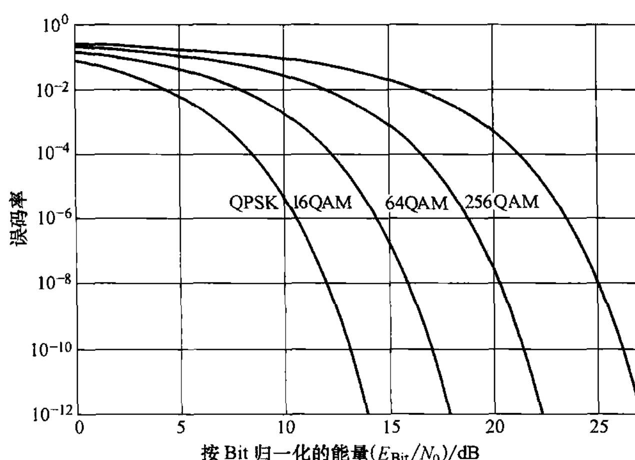


图3-75　QAM方式下的误码率

因此 P_S/P_N 的值最小必须是13.7dB。对 P_N 来说，可以使用式（1-20），因此，

$$P_N = N_0 B = kT\frac{R}{2}(1+r)$$

$$= 1.38\times10^{-23}\times293\times\frac{2\times10^6}{2}\times1.35\text{W} = 5.46\times10^{-15}\text{W} \hat{=} -142.6\text{dBW}$$

因此在接收器中的 P_S 最小值应为 -128.9dBW $= -98.9$dBm。设计时应预留几个dB，因为每段线路都有衰减（第5章将讨论这个问题），由此可以推算出发送器输出端的信号功率。

传输系统的各项指标可以根据不同的方法来确定，例如可以按照下面的方法进行：

1）从应用出发，确定数据传输速率 R 和误码率。

2）确定可用带宽 B_u，但这只是物理层的数据，也要遵照规定和标准的要求。

3）确定接收端的噪声功率 P_N 及信道的衰减 A_K。

4）考虑到技术上可行的发送功率及 A_K，计算接收端的信噪比，由此确定调制方法和传输符号编码个数 M。这点应该仔细考虑信道编码的情况。

5）然后再检查一下所选择的方法是否能保证所要求的数据传输速率。如果不行，应该提高带宽增大信噪比。

根据下面的判据可以对不同的调制方法进行比较评价：式（3-61）给出的频谱效率、图3-75给出的误码率、最大信号功率与平均信号功率的比（这对放大器的动态性和线性要求很重要，它会影响电路的技术指标）、调制和解调方法的复杂性（它决定设备的价格）。

3.4.6　正交频分复用

根据上面所述的方法确定传输链路后，得到如下的结果，采用16QAM，用1MHz带宽带传输4MBit/s的数据（为简化起见这里假设 $r=0$）。

所谓的多载波方法的主要思想是不像上面所述的方法那样利用一个大带宽的信道，

而是用许多小带宽的信道，即子信道来完成相同的任务。它涉及的是图 1-38 所讨论的向下复用（Down-Multiplexing）的情况。上面的系统可以采用例如下面的方法来代替。

1）10 个子信道，每个子信道的带宽为 100kHz 并使用 16QAM。这样每个子信道可传输 400kBit/s，总共仍可达到 4MBit/s。

2）10 个子信道，每个子信道的带宽仍为 100kHz，但其中的 5 个信道采用 64QAM，每个信道的速率为 600kBit/s，其余信道采用 QPSK，每个信道的速率为 200kBit/s。这样总的数据传输速率仍为 4MBit/s。

从带宽调制到多载波系统的过渡也可以理解成是从时分复用到频分复用的转变。

多载波系统与单个带宽调制相比在下面两点上很有优势：多径接收和选择性衰减（→应用上面的方法 1））、信噪比 SR_K 随频率发生变化（→应用方法 2））。

（1）多径接收和选择性衰减

短波及 VHF/UHF 波段的陆地无线传输经常会遇到多径接收的问题。接收器会收到直接传来的无线信号，也会收到经过山岗、建筑物等反射有了一定的延迟和衰减了的无线信号。例如电视接收中的重影就是由多径接收引起的后果。

例如上面例子中的 4MBit/s、16QAM 系统的符号周期为 1μs。无线电波在这段时间内可传 300m，直接接收和经过反射接收信号的距离只有 300m，这样在接收端，经过延迟的符号和直接发过来的下一个符号就会同时接收到。因此多径接收会引起很大的码间干扰（ISI）及选择性衰减（窄带破坏性干扰）。

改进措施很明显：增大符号周期。每个符号的编码个数可以不改变，它与信噪比有关。不过相应地必须降低数据传输速率及带宽，但这可为其他信道赢得空间。甚至可以通过所谓的保护间隔（guard interval）来人为地延长符号，这样虽然降低了信道的利用率，但却有可能完全消除码间干扰。

多径接收会涉及到线性失真，原则上这可以通过图 1-60 所示的均衡器来补偿。问题是在宽带传输的情况下，符号周期与时延相比较短，这要求均衡器的阶数不切实际地高。相反 OFDM 的窄带子信道允许使用较简单的均衡器（如果必须使用的话）。使用 OFDM，调制器和解调器中的均衡器的复杂性需要重新考虑。但最主要的还是有一个比较便宜的解决方案，因为调制器和解调器可以很有效地利用 FFT 来实现。

如果多载波系统可以处理多径接收的问题，那么它也可以实现同播网络。这样通过多个发送相同节目的发射器可以实现较大的覆盖范围，此时多个发射器之间必须进行同步，并且使用相同的频率工作。这样可以很经济地利用频谱资源实现无线广播，新的数字广播和电视系统（DAB 和 DVB）就使用这种技术[Nee00]。

（2）随频率变化的信噪比 SR_K

铜线电缆受集肤效应的影响会有衰减，它随频率的升高而加大。因此宽带传输比窄带传输受的影响更大。一种可能的应对办法是使用预加重/去加重（Preemphase/De-emphase）的方法（参见 1.3.1 节）。多载波系统同样也是一种对策，系统中每个子信道的传输符号编码个数随该信道的信噪比而变化。

在点到点的双向传输（所有的 ARQ 系统）中，接收器可以不断地向发送器报告它现在的接收状况。这样调制方式就可以动态地进行调整（即自适应系统）。

这样的一个应用例子就是ADSL（Asymmetrical Digital Subscriber Loop，参见5.6.4节）。这样双绞铜线的传输速率可以达到8MBit/s。

多载波方法的基础是式（1-17）所给出的信道能力。那里假设 P_S 和 P_N 是常数，但实际不一定是这样的，例如受干扰的发送器可能只是影响宽带信道的一小部分。式（1-17）给出的只是一种通用的情况：

$$C = \int_{f_u}^{f_o} \log_2\left(1 + \frac{R_S(f)}{P_N(f)}\right) df \tag{3-62}$$

当然式（3-62）有些不实际。一种近似的方法是将总的带宽分成多个段，而每段的信噪比都是常数。因此式（3-62）可以改写成

$$C = \sum_{k=0}^{N-1} B_k \log_2\left(\frac{P_{S_k}}{P_{N_k}}\right) \tag{3-63}$$

现在我们来看看多载波方法的实现。每个子信道都采用QAM，它使用如图3-72所示的电路来实现。利用数学方法可以紧凑地描述载波方法，这在式（3-59）中已经使用过。数字信息信号首先用解复用的办法分成复数信号部分 $s_k(t)$，然后使用载频 ω_k 进行正交调制，最后将所有子信道的信号加在一起（频分复用），即得到多载波信号：

$$s_{\text{Multicarriet}}(t) = \sum_{k=0}^{N-1} s_k(t)\, e^{j\omega_k t} \tag{3-64}$$

当然这个方法很复杂。但是，有一种特殊情况可以很方便的实现：载波之间相互正交。这就是OFDM（Orthogonal Frequency Division Multiplex）方法。当满足下面条件时我们称载波是正交的：

$$\omega_k = 2\pi k f_0 \tag{3-65}$$

f_0 是基频，它根据符号周期来调节。每个QAM信号的频谱都是 $\sin(x)/x$ 曲线，曲线最大值左侧和右侧的零点之间的间隔 Δf 满足：

$$\Delta f = 2/T_{\text{symbol}} \tag{3-66}$$

T_{symbol} 是子信道的符号周期，包括保护间隔。OFDM方法中所有子信道的符号周期都是相同的，在OFDM中式（3-65）中的 f_0 通常设成式（3-66）中值的一半，即 $\Delta f/2 = 1/T_{\text{symbol}}$。因此，式（3-64）变成如下：

$$s_{\text{OFDM}}(t) = \sum_{k=0}^{N-1} s_k(t)\, e^{j2\pi k \frac{t}{T_{\text{symbol}}}} \tag{3-67}$$

图3-76所示为正交频率的选择：在频率 kf_0 处只有一个 $\sin(x)/x$ 曲线的值不为零。

因为我们讨论的是数字信道，所以时间变量 t 是离散的。因此式（3-67）与离散傅里叶逆变换（IDFT）是一致的，而后者利用FFT（Fast Fourier Transform）有一种很有效、很快速的解法。它通过算术逻辑单元采用数字的办法即可实现，即频率无需转换到UHF频段。因此可将调制进行分解，即分解成一个复杂的但可用数字方法实现的

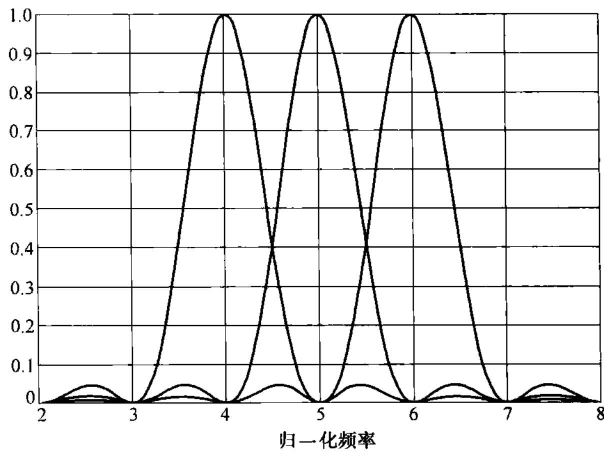


图 3-76　3 个正交 QAM 信号的功率谱

预置级，后接一个 SSB 发送器实现频率转换。图 3-77 给出的是 OFDM 发送器的原理图。

图 3-77 中，数据流中时间上相邻的比特被放入 OFDM 信号的同一个子信道中。这是不希望的，特别是当数据流采用信道编码被保护之后。信道编码是很简单的，例如当一个 100Bit 的数据流中的第 20、54 和 83Bit 发生错误时，可用第 29、30 和 31Bit 来代替。但两种情况下的误码率（BER）都是相同的。显然在无线传输中，一个符号受相邻符号的干扰而发生错误的概率要比受相差很远的符号的干扰而发生错误的概率大。因此可在 IDFT 之前采用时间交织器或在 IDFT 之后采用频率交织器对数据流进行重组。这就是所谓的编码正交频分复用（Coded Orthogonal Frequency Division Multiplex, COFDM）。注意这里信道编码和调制相互之间也有关联。

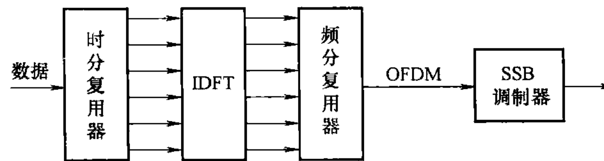


图 3-77　OFDM 调制器

3.5 多级调制

1. 两种方法相互交织

调制的任务是：复用（集中）和信道适配（频率搬移和干扰保护）。

因此可将它们分成两类独立的调制方法，它们之间会有交织，如图 3-78 所示。

下面是几个例子，a）和 b）分别表示上面的两个任务，即复用和信道适配。

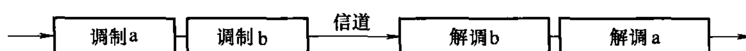


图 3-78　调制类型的交互使用

a) SSB　b) FM　　例：经过复用（FDM）的 SSB 信号采用 FM 进行传输。模拟电话通过无线链路传输时就采用这种方法。

a) PCM　b) PSK　例：TDM 复用采用 PSK 无线链路进行传输。

a) RFM　b) ASK　例：RFM 信号控制 LED 的开和关（光传输时）。使用 RFM，LED 的非线性特性对传输就不会产生影响。而光信号本身采用 ASK 进行调制。

2. 同时对两个载波参数进行调制

这种方法的意义在于可以传递额外的信息，它付出的代价是抗干扰性会受到影响。

(1) AM 与 FM 相互结合

AM 的调制深度会因此收到影响，不能做得很大。

(2) PAM 与 PFM、PWM、PPM 或 PCM 相结合

PAM 的信息在幅度轴上，而其他调制方法的信息都在时间轴上。利用 PCM 链路中的某一段作为业务信道就使用这种方法。

(3) ASK 与 PSK 相结合

已经讨论过的 QAM 就是这样做的。

3.6　调制类型的命名

众多可能的调制方法以及已经得到实际应用的调制方法可以按照统一的方式进行分类。下面所介绍的分类方法是 1979 年引入的，引入的目的是用来替代旧的，不再能满足需求的分类系统。

但是并不是说这个分类系统就没有问题。因为不仅发送的信号需要描述，它的源类型也需要描述，因此有时对信号分类存在多种可能性。

每个类由 3～5 个字符，按照 BZBBB 的方式进行描述。这里 B 指字母，Z 指数字。第四位和第五位是可选的，它是对信号更进一步的刻画。表 3-6 到表 3-10 给出分类的描述方法。

表 3-6　调制分类的第一位（字母）

	第一位	含义：主载波的调制
	N	没有调制的载波
AM：	A	DSB(普通的双边带 AM)
	H	SSB，全载波
	R	SSB，减少的(可变的)载波
	J	SSB，抑制载波(SSSC)

（续）

	第一位	含义:主载波的调制
	B	ISB(独立边带)
	C	VSB(残留边带 AM)
WM	F	FM
	G	PM
AM + WM	D	
脉冲调制	P	没调制的脉冲序列
	K	PAM
	L	PWM
	M	PPM
	Q	带有共存载波的 WM
	V,W,X	其他情况

表 3-7　调制分类的第二位（数字）

第二位	含义:信息信号的类型
0	未被调制信号
1	信息信号使用单信道,经过量化或数字化,未经过预调制(TDM 信号除外!)
2	信息信号使用单信道,经过量化或数字化,经过预调制(TDM 信号除外!)
3	信息信号使用单信道,是模拟信号
7	信息信号使用多信道,经过量化或数字化
9	模拟信号和经过量化或数字化的信号相结合,使用单信道或多信道
X	其他情况

表 3-8　调制分类的第三位（数字）

第三位	含义:信息传输类型
N	无信息
A	接听电报(Morse,摩尔斯式电码)
B	自动接收电报(电传打字机)
C	传真
D	数据、遥测、遥控
E	电话和声音广播
F	电视
W	上面情况的组合
X	其他情况

表 3-9　调制分类的第四位（数字）

第四位	含义:信号的详细描述
A	符号持续时间不同的二元码
B	符号持续时间相同的二元码,无纠错
C	符号持续时间相同的二元码,有纠错

（续）

第四位	含义:信号的详细描述
D	四元码
E	多元码
F	描述文字的四元码
G	与无线广播质量相当的单声道音频
H	与无线广播质量相当的立体声或双声道音频
J	到达商用质量的音频信号(不包括下面的K类与L类)
K	到达商用质量的音频信号,有频率转换或划分子信道
L	到达商用质量的音频信号,有单独的导频来控制调制(Aussteuerung)
M	单色图像
N	彩色图像
W	上面情况的组合
X	其他情况

表3-10 调制分类的第五位（数字）

第五位	含义:复用的类型
N	未复用
C	码分复用(CDM),包括扩频技术
F	频分复用(FDM)
T	时分复用(TDM)
W	频分复用和时分复用结合
X	其他情况

最后还需给出已调信号的带宽，先看几个定义：

1）占用带宽：范围的界限大多数情况下根据功率相对总平均功率下降了0.5%（-23dB）（比较图1-42）时来确定。

2）必要带宽：占用带宽中足够用于信息传输的那一部分。

3）分配带宽：必要带宽加上频率允许偏差绝对值的两倍。

带宽用三位数字和一个字母给出，字母用来代替十进制数的小数点并同时给出数值的单位。字母可以是H（Hz）、K（kHz）、M（MHz）及G（GHz）。数值四舍五入只保留3位，表3-11给出几个例子。

表3-11 频率描述实例

频 率	描述方法	频 率	描述方法
0.002Hz	H002	180.5kHz	180K
0.1Hz	H100	10MHz	10M0
12.5kHz	12K5	5.65GHz	5G65
180.4kHz	180K		

第4章　编　　码

编码指由一个信号构造成另外一个信号，即为实现下面的目的而进行的信号转换：

1）减少数据量　　　　　　→信源编码，4.1 节

2）保护私密性　　　　　　→加密，4.2 节

3）信道适配（干扰保护）　→信道编码，4.3 节

不论是在发送端进行的信号转换还是在接收端进行的反变换都是很复杂的操作，因此信源编码是最主要的，实际中信道编码和加密只应用于数字信号。

线路编码的目的是获得时钟并且根据应用的不同有可能还包括消除直流分量。尽管后者也可认为是信道适配，但并不将线路编码归结成信道编码。因此已经将它单独在 2.4 节中介绍过了。

信源编码的目的是减少数据量，根据图 1-12 可以通过减少冗余和/或减少不相关的办法来实现。而信道编码却刚好相反，它增加冗余，但目的是改善传输的可靠性。从信源编码转到信道编码绝不是两个互补的步骤，因为去掉的原始冗余很少会用来检错。图 4-1 给出的是相连的不同编码模块，箭头的粗细代表相应的数据传输速率的大小。

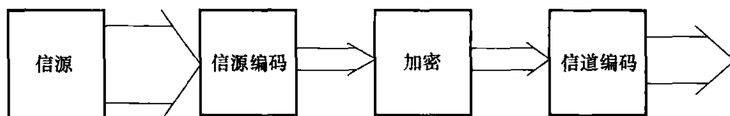


图 4-1　不同编码的顺序（箭头的粗细代表相应的数据速率的大小）

通常将信源编码归为用户侧，而将信道编码归为传输侧，如图 4-2 所示。对加密

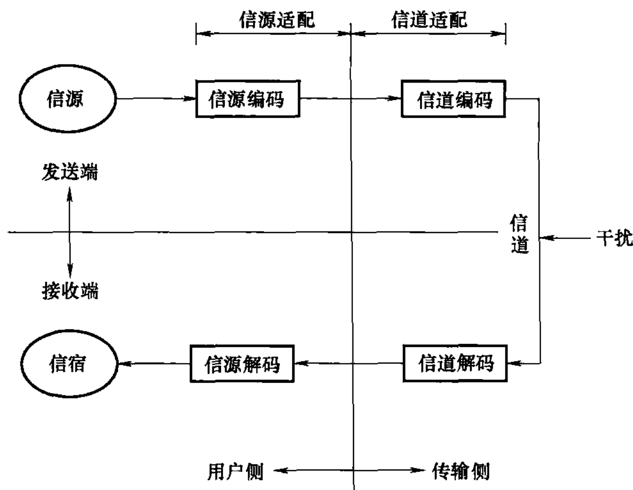


图 4-2　信息传输系统用户侧和传输侧一种可能的划分方法

传输而言，加密是在信源编码和信道编码之间进行。在信源编码之前进行加密有很大的数据量要进行处理，并且加密后的信源编码根本不能对加密的信息进行解释。对不进行加密的电话网来说，加密可由用户自己来完成。而对加密的网络，如军用网络，加密属于传输的一部分。后者经常使用所谓的复用编码设备将整个 TDM 信号进行加密。

用户侧和传输侧的区分由网络运营状况来确定，属于管理方面的内容，但这不应与传输信道的界限混淆，传输信道描述的是技术系统理论方面的内容。

4.1 信源编码

信源编码有很大的经济价值，它减少了必要的带宽和/或传输时间以及内存需求；因此传输费用也会降低。当数据量较大时，如活动图像（视频）信源编码显得尤为重要。

当然信源编码也有缺点：系统的复杂性（需要的计算能力）会提高，即传输费用可能会被设备本身的费用来替代。但设备会因微电子技术的迅猛发展而迅速下降，因此进行数据压缩还是很值得的。因为编码和解码都需要一定的时间，所以对实时应用进行信源编码可能会产生一些问题。

4.1.1 减少冗余（算术压缩）

减少冗余和减少不相关是不同的，它是一个可逆的过程。编码相当于根据信源的符号明确地构造出传输代码的符号，如图 1-48 所示。减少冗余不会改变有效信息的内容。

减少冗余的方法主要分成以下几类：霍夫曼编码、预编码、行程编码、字典编码。

我们首先要仔细分析一下，信息量如何理解。在 1.1.4 节中我们给出了一个直觉的定义，现在我们对它用数学的方法进行描述。信息我们认为是通过接收一系列符号，消除这些符号中的不确定性而最终得到的内容。这里信息的量与它出现的概率有关。假设信源 N 发出不同的符号 S_i，每个符号出现的概率是 p_i，这样每个符号的信息量为

$$H(S_i) = \log_2\left(\frac{1}{p_i}\right) = -\log_2(p_i) \tag{4-1}$$

符号的平均信息量被称为信源的熵：

$$H = \sum_{i=1}^{N} p_i \cdot H(S_i) = \sum_{i=1}^{N} p_i \cdot \log_2\left(\frac{1}{p_i}\right) = -\sum_{i=1}^{N} p_i \cdot \log_2(p_i)\ [\text{bit/符号}] \tag{4-2}$$

一个很重要的特殊情况是二进制信源，它有两种符号且两种符号的概率分别为 p 和 $1\text{-}p$。因此二进制信源熵函数为

$$h(p) = p\log_2\left(\frac{1}{p}\right) + (1-p)\log_2\left(\frac{1}{1-p}\right) \tag{4-3}$$

图 4-3 所示为 $h(p)$ 的曲线。

在式（4-3）中，当 $p=1$ 时，信源发出的肯定都是符号 1 而没有符号 2。对接收者来说，这一点好奇也没有，因此相应的信息量是 0。最大熵 $H_{max}=H_0$ 发生在 $p=0.5$ 时，此时是对称的二进制信源。

可以证明，当信源有 N 个符号时，最大熵 H_0 出现在各个符号的概率相同的时候。此时式（4-2）中概率 p_i 的值相等，都为 $1/N$，因此由式（4-2）可得

$$H_0 = \sum_{i=1}^{N} \frac{1}{N}\log_2(N) = N\frac{1}{N}\log_2(N) = \log_2(N) \tag{4-4}$$

这里 H_0 称为决策量，它由信源字符的基数决定。

最大熵和平均熵之间的差称为冗余 R。冗余是衡量一个信号的可预测度的手段。

$$R = H_0 - H = \log_2(N) + \sum_{i=1}^{N} p_i\log_2(p_i) \tag{4-5}$$

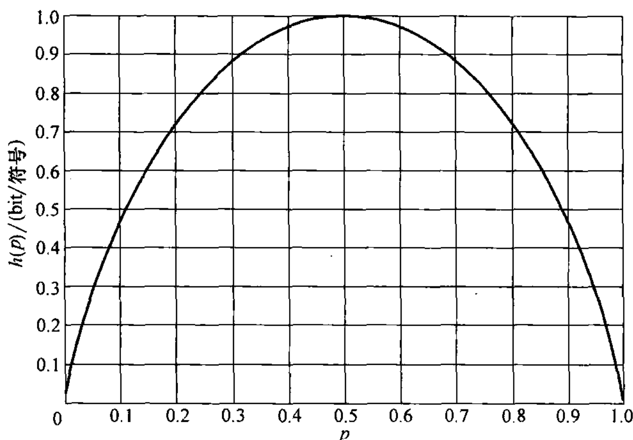


图 4-3 二进制信源的平均信息量（熵）

作为例子我们来看一个信源：4 个符号 A、B、C、D 出现的概率分别为 0.5、0.25、0.125 和 0.125。那么熵的数量为

$H = 0.5\log_2(2) + 0.25\log_2(4) + 0.125\log_2(8) + 0.125\log_2(8) = 1.75$bit/符号

相应的符号流如下：

AAAC ACAB ABBA BADA ACBD AABD…

如果所有符号出现的概率相等，那么最大熵 $H_0 = 2$bit/符号，此时冗余值为 0.25 bit/符号。

现在我们看看如何将信源符号变成二元码。可以按照下面的方式分配：

A 00

B 01

C 10

D 11

上面例子中的符号流有 24 个符号，每个符号的平均信息熵为 1.75bit/符号，因此总的信息量为 42bit。上面建议的编码给每个符号分配 2Bit（大写 B!），这样数据量总共为 48Bit，而它的信息量为 42bit，因此冗余是这样的：每个信源符号 0.25bit，这样每个 Bit 的冗余为 0.125bit，48Bit 的冗余共为 6bit。

一个较好的编码方法是使用决策树，首先询问信源符号是否为 A。否定的情况为 50%，因此继续询问它是否为 B。若答案是否定的则继续询问是否为 C。我们将肯定回答编码为 1，否定回答编码为 0，因此分配方案如下：

A　1

B　01

C　001

D　000

这样传输上面的符号流需要 $12\times1+6\times2+3\times3+3\times3=42$Bit。这样数据量与信息量相等，冗余被消除了。此时 0 的个数与 1 的个数相等。但实际情况并不总像所举例子这样完美。

减少冗余的最基本的想法是：信源符号按照它们的频率及出现的概率排序，传输代码由不同长度的符号（数据字）组成，最短的数据字分配给最常出现的信源符号。根据算法的不同，有香农码（Shannon Code）、费诺码（Fano Code）、霍夫曼码（Huffman Code）。

霍夫曼码比费诺码复杂，但效果更好，香农码在实际中不会应用。

如果所有信源符号的概率相同，数据流不能够被压缩，因为它本身就不包含任何冗余。离散随机过程就可以证明这点。如果可以将符号序列用算术的方法进行压缩，即它就不会包含随机序列。

减少冗余可以理解为将数据流中存在的任何一种结构作为先验信息，并在传输之前将它消灭掉。预编码也是这样，我们在差分 PCM 中已经遇到过了。

压扩器的瞬时值（例如 PCM 传输中 A 率的形式）也会影响冗余的减少。因为它改变了信源符号（这种情况下即是话音信号的瞬时值）的概率分布，即从图 3-40 所示的分布变成图 3.39 所示的等概率分布。这样做的收获是减少了量化噪声，若采用线性 PCM 编码则需要较高的传输速率才能实现。

减少冗余也可以理解成是去相关，因为压缩之后信号本身的特性与随机信号完全相同。

顺便提一下，自然科学作为一个整体不外乎就是寻找一个压缩算法。即研究一个被观察的序列如何能用一个很紧凑的公式来表达。

下面的例子深化减少冗余的主要原理。表 4-1 给出了一个信源，它有 3 个符号 A、B 和 C，每个符号出现的概率各不相同。同时该表还给出了 3 种编码方式。

表 4-1　不同信源编码的例子

符号 S	$p(S)$	编码 1	编码 2	编码 3
A	0.6	0	0	0
B	0.3	01	10	10
C	0.1	10	110	11

表 4-1 给出的编码方法并不都是经过精心选择的。第一种编码方式不能够解码，因为数据流…0010…既可以翻译成 AAC 也可以翻译成 ABA。从这点上来看，编码 2 和

编码3较好：编码2是所谓的逗号码，即每个码字都有一个明确的结尾，这里是0。编码2和编码3都满足所谓的前缀特性：没有码字与其他码字的开始部分相同。这样的编码也称为自同步。前缀特性在图4-4所示的所谓的编码树中很容易理解。

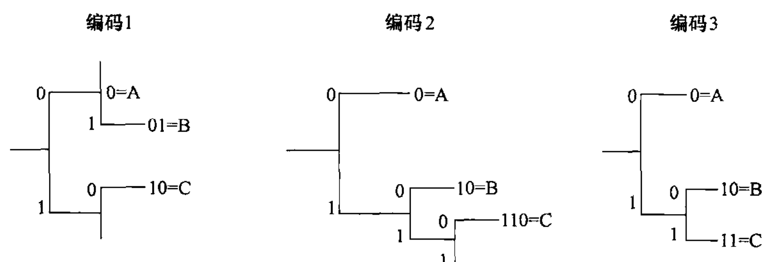


图4-4 表4-1的编码树（编码3是霍夫曼编码）

二元码的编码树由二叉分支组成，每个二叉分支有两种可能。如果选择上面的路径，就意味着0，下面的路径就是1。当然相反的分配方案也是可行的。自同步编码的码字始终是在分支的末端，从不在分叉上。

此外，摩尔斯电码也是一种信源编码，英语（德语也同样）中经常出现的字母e被分配给最短的码字点。摩尔斯电码不是自同步的，因此它实际使用3种符号，即点、划和停顿。

上面是信源编码的基本原理。下面还有一些技巧：

1）一段文字由字母组成，每个字母出现的概率不同。但是实际上经常出现的是一些固定的组合，例如在一段德语文字中ei、en、er、ie、ch、ck、sch、qu等字母组合经常出现。人们称之为有记忆的离散信息源。当引入所谓的超级符号时，可对多个字母同时编码（即马尔可夫编码），这样压缩得会更好。甚至可以将整个词进行编码，整个句子甚至整本书也可以这样。从原理上说可以将整体上的冗余消除掉，但是这样做复杂性很高。具有这种属性的编码称之为最优码。

德语书写语言的决策量H_0的值为$\log_2(26)=4.7$bit/符号。考虑到每个字母出现的频率，有熵$H_1=4.1$bit/字母。如果使用经常出现的字母组合和音节，可以得出$H_3=2.8$bit/符号。如果将整个词进行编码，可得$H_6=2.0$bit/符号，对整个句子编码，那么$H_\infty=1.6$bit/字母。

2）上述编码方法的结果是代码的长度是可变的。如果偶然不经常出现的字符出现多次，会因此出现较长的码字。信道的负载就会超出平均值。如果使用中间存储器（缓冲），就可以避免这个可以容忍的延迟。

3）对于传输来说，不等长的码字是有缺点的，特别是当一个数据流与其他的信号复用在一起时（TDM）。因为编码是自同步的，数据流在传输之前可能被分成等长的字，它们在接收端又被重新接在一起，由信源解码器先正确地恢复成不等长的字，然后再进行解码。

这些编码的缺点是信源的特性必须是已知的。因此必须将编码分成两个、可能是比较耗时的阶段来完成，首先确定信源的特性，然后再进行编码（可适配编码）。

数据传输速率由平均码长来决定。我们设码长为 W，那么它的平均值就是期望值 $E[W]$。当然 $E[W]$ 最小应该等于信源符号的平均信息量（即信源的熵），这样才不会有信息丢失。

$$H \leqslant E[W]$$

现在的问题是是否能给出 $E[W]$ 的上界。可以证明，自同步的最优码满足下面的关系：

$$H \leqslant E[W] < H+1 \tag{4-6}$$

相反 Lempel-Ziv 算法与信源特性无关。它使用相反的方法，将信源流不等长的字转变成长度为固定值（缩短）的码字。这种算法在对数据文件在传输或存储前进行压缩方面有很广泛的应用。这种算法也包含在计算机的操作系统中。

Lempel-Ziv 算法也是字典编码的最著名的代表。

最后我们还得提一下行程编码（run length coding）。这种编码方法在图像和传真传输方面应用得很成功。一个 Run 是一串 0 Bit 或 1 Bit。例如序列... 001111000... 包含一个长度为 2 的 0-Run，一个长度为 4 的 1-Run 以及一个长度为 3 的 0-Run。因为通常情况下图像中相邻的点是相同的，经常会出现较长的 Run。0-Run 与 1-Run 的改变由清晰度来确定，Run 根本不用进行编码。例如上面的序列可以用十进制码的... 243... 来传输。尽管这种方法同步起来很复杂，但它已经将压缩考虑进去了。但如果对随机图形编码，那么信源编码后的数据量有可能更大。

4.1.2　减少不相关（减少熵）

与减少冗余不同，减少不相关是一个不可逆的过程，即信息丢失是不可恢复的，也可以用质量丢失来表达。为了使信息丢失不会带来较大的影响，针对具体的应用，需要与信宿进行协商：哪些信息是相关的，哪些是不相关的或无关紧要的。

因为编码过程可能会丢失一些信息，所以人们也称之为有损数据压缩或减少熵。与此相反减少冗余是无损压缩。

减少冗余考虑的是信源的特性。 减少不相关考虑的是信宿的需求。

由于通常情况下信宿是人类本身，因此在音乐、图像和视频传输中减少不相关是一个很重要的话题。

例如：

1）量化属于减少不相关。如果在 PCM 信号中 LSB 的每个数据字都设为零，那么它们是个常数，即没有任何信息，所以也就没有必要传输了。字宽减少 1Bit 的 A/D 转换器也有相同的减少数据的效果。CD 中 16Bit 的解析率被认为是足够了；人们对 PCM 话音传输中使用 8Bit 产生的效果已经很满足了。

2）电视：当每秒钟连续播放 20 幅或更多的静止图像时，人的眼睛就认为是活动图像了。因此在电视中每秒钟传输 25 幅图像就足够了，而无需传输 100 幅或更多。但当室内的昆虫看到电视时，它会认为这是每秒钟 25 幅独立的图像。

3）彩色电视：人眼视网膜柱细胞的作用是识别亮度，锥细胞的作用是识别色度。由于锥细胞对角度的解析不如柱细胞好，所以人眼对图像清晰度的感觉主要来自于对亮度的敏感度。这点在小孩子们的绘画中已经用到，小孩子们在事先画好的黑白轮廓中很笨拙地图上颜色，并且还需要家长的鼓励。在彩色电视中也用到是个事实，亮度信息使用大约5MHz的带宽进行传输，而颜色信息仅使用1.3MHz的带宽就可以了。

4）音频带宽：人在年轻时最多能听到20kHz以内的声音。因此音频信号的带宽及相应的采样频率最多是20kHz和44.1kHz。人类本身能够发生的声音最多是几kHz（不使用任何乐器），因此3.4kHz的带宽传输话音就足够了，即采样频率为8kHz。

比较困难的是减少不相关和减少冗余结合在一起使用。这通常需要不同领域的人共同努力，因此需要较宽的知识面。

1）数据信号只能使用减少冗余的办法，因为它要求不能有任何数据丢失，参见4.1.1节。

2）话音信号使用减少冗余和减少不相关相结合的办法，参见4.1.3节。

3）音频信号（音乐）具有很大的多样性，因此主要使用减少不相关的办法，参见4.1.4节。

4）图像信号（静止图像）出于相同的考虑主要是去掉不相关的内容，参见4.1.5节。

5）视频信号（活动图像）采用减少冗余和减少不相关相结合的办法，参见4.1.6节。

4.1.3 话音信号压缩

下面的几种可能性在前面已经提到过了：减少带宽（减少不相关）、相对较粗粒度的量化（减少不相关）、瞬时值压扩（减少冗余）、预编码（DPCM、DM、ADM、Sigma-DM）（减少冗余）。

减少冗余的方法也称为波形编码（waveform coder）。它与信号的类型无关，因此具有很好的鲁棒性并且得到了广泛的应用。

话音编码器（voice coder）的工作方式完全不同，它使用一系列滤波器（实际中采用窗口大小为16～20ms的短时间FFT来实现）在整个频率范围内构造一系列的子频段，并将能量分配到每个子频段上。数值将在这些子频段上被传输，这样在接收端才能执行相反的过程对信号进行近似。在这里利用了人类的听觉对相位误差并不敏感的原理，参见表1-6。

最有效的声音数据压缩是通过所谓的分析综合方法实现的。这里人们想办法对信号的产生过程进行近似，而不是信号本身。人们对声源进行建模，并给出模型的参数。如果对声音进行限制，那么一个模型就足够了，问题就简化成确定模型的参数。这种方法称为LPC（Linear Predictive Coding，线性预测编码）方法，我们将进一步分析它。

图4-5所示为确定参数的一般过程。源系统和模型使用相同的输入信号。优化的作用是确定参数，使输出信号 $y(t)$ 和 $y'(t)$ 的差尽可能小。

如果模型结构与实际情况较符合，人们可以在模型内规定在系统中不容易得到的值的大小。自动控制技术就是这样做的（状态-观察者）。

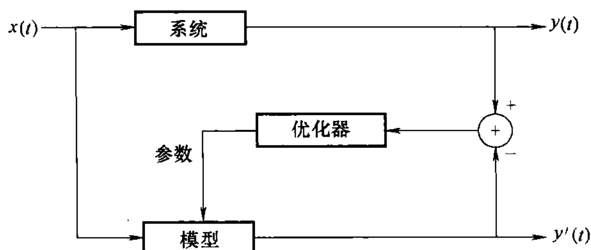


图 4-5　参数确定

如果已知激励信号，模型不需要输入信号或者说模型可以自行产生输入信号。因此 $y'(t)$ 仅与参数有关。因此信号 $y(t)$ 可以用信号 $y'(t)$ 来近似。而 $y'(t)$ 可以简洁地通过参数来描述。使用这种方法就可以实现数据压缩。

人类发出的元音由来自舌头的气流来控制。确定声音大小的参数是口腔、双唇等的瞬时形状。元音发声可以通过线性 Allpolfilter 来建模，这里 Allpolfilter 是指没有零点，每 16～20ms 更新一次参数的滤波器。当采样频率为 8kHz 时，有 128～160 个采样值。元音发声的激励是不能测量到的，在模型中使用两个不同的信号发生器来代替：一个是噪声发生器，主要用于形成非浊音话音信号；另一个是周期脉冲发生器，主要用于形成浊音话音信号。脉冲的循环速率与所谓的定音频率相符，即表现为话音信号的音高。图 4-6 所示为话音产生器的模型。当参数被正确地确定时，输出信号 $y'(t)$ 与原始的话音信号很相近。$y'(t)$ 可以根据使用的信号源（开关设置）、定音频率及参数的值来描述。

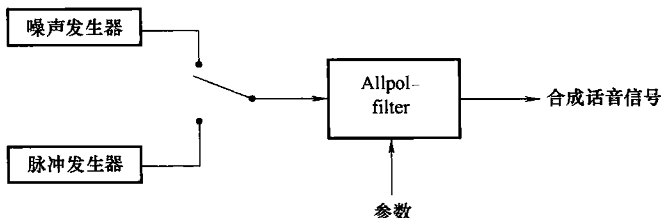


图 4-6　对人类元音发声的建模

在话音传输中需要节省带宽，为此人们并不传输话音信号 $y(t)$ 或它的近似 $y'(t)$，而是传输参数的值（5～10 组值，每组用 10Bit）、使用的源（1Bit）、定音频率（6Bit）及声音的大小（每种压缩特性 5Bit）。在接收端使用相同的模型并提供相同的输入参数，这样在接收端就会产生信号 $y'(t)$，如图 4-7 所示。

参数的计算采用最小二次方误差的办法，即对线性方程组进行求解。使用 FFT 确定正确的源（噪声、脉冲及定音），及在频域内寻找等距离的峰值也包括其中。

传输参数比直接传输 $y(t)$ 或 $y'(t)$ 时信息量要小，而且接收端能够完全恢复 $y'(t)$。这与信息论并不矛盾，原因是在模型中已经包含了许多话音信息，并不需要传输这些先验信息。

一个小例子会使这个概念更加清晰：一首钢琴曲可以使用传声器来捕获、数字化

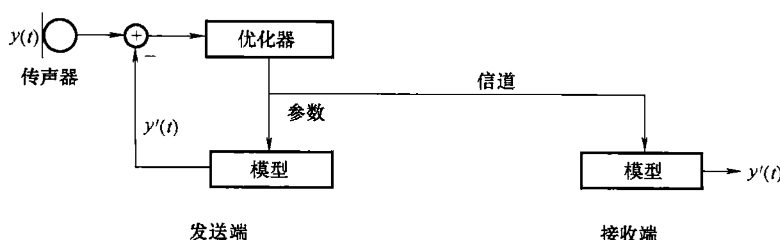


图 4-7 用 LPC 方法进行话音数据压缩

和传输。这么大的信息量需要使用很大的带宽来传输。但是当人们知道它是一首钢琴曲（先验信息），那么在发送端可以用按下的琴键（参数）来代替声音。这样传输的就只是一系列的琴键号（与曲谱相同!），在接收端可以使用另一架钢琴（模型）按下相同的琴键即可。这样就可以听到与原始声音相同的钢琴曲，只传输了很少的信息。节省的信息是与钢琴有关的知识。如果发送端和接收端的模型不一致（例如发送端是钢琴，而接收端是拨弦古钢琴），那么肯定会出现错误。如果模型相同但却不合适，那么 $y'(t)$ 也会与 $y(t)$ 相差很远。合适的模型通常很复杂，所以有可能不能实时地确定参数。在实际中会采取许多折中方法。当有一点先验信息时，相应地会有许多参数需要估计。极特殊的情况下，参数可能与原始信号有相同的信息量。

CD 不需要使用 LPC 方法，因为人们不能找到一个实用的模型来表示众多的声音种类。人们将这种方法仅限制在话音上，它对话音合成很有帮助。

LPC 方法也用在 GSM 移动电话中，这里的采样频率是 8kHz，解析率是 13Bit，所以数据传输速率是 103kBit/s。使用 LPC 可以将它减少到 13kBit/s，并且能够保证很好的话音质量。后续的信道编码又将它扩展到 22.8kBit/s。另外一个例子是数据传输速率为 2.4kBit/s 的使用卫星的 Inmarsat 电话系统。

上面描述的 LPC 方法还有许多扩展，部分也已经得到标准化。表 4-2 概括了这些话音编码方法。

表 4-2 不同的话音编码标准

编码标准	描述		数据传输速率/(kBit/s)
G.711	PCM	(脉冲编码调制)	64
G.722	SB-ADPCM	(子边带自适配差分 PCM)	48,56,64
G.728	LD-CELP	(低时延代码激励线性预测)	16
G.729	CS-ACELP	(共轭结构 ACELP)	8
G.723.1	ACELP	(代数代码激励线性预测)	5.3,6.3
	MELP	(混合激励线性预测)	2.4

话音编码和波形编码很少对信源做假设，即使用的先验信息较少因此提供的压缩增益也较小。在相同的话音质量下，话音编码的数据传输速率比 LPC 要高很多。尽管如此，它还有存在的意义：它使存储所有音素的特征值（音素的基本单元）成为可能。基于它可以构成所有词，使用话音编码音素库可以直接将文字转换成话音。

显然，实现上述的概念需要多级且要求很苛刻的信号处理步骤。因为每一步都会给信号增加一些误差，那么到结尾信号可能不可用。但如果能在处理中增强信号对噪声的免疫力，那么这些方法才是比较实际的。这里的增强免疫力是指量化，即采用数字系统来实现这些概念。

如果对不同的双向话音信道进行复用，那么传输系统的利用率只能达到40%左右。这是因为绝大多数情况下只有一个人在说话并且存在话音停顿。使用动态信道分配可以改善比较昂贵的信道利用率，它的做法是当一个话音信道中的话音有较长的停顿时，将该信道分配给其他的用户使用（减少不相关）。在接收端使用噪声来模拟该用户的存在。这样做对话音理解的影响是在开始说话时会有部分词被截掉现象。改进办法是在说话前清清嗓子或咳嗽一下，这样可以重新抢回信道。由于动态信道分配增加了管理上的复杂性，但是这样做比像卫星这样只有一半利用率的方式还是便宜的。这种技术叫做DSI（Digital Speech Interpolation），在海底电缆中也使用这种技术并称之为TASI（Time Assignment Speech Interpolation）。使用这种方法在100个全双工话音信道中可以实现250个对话。在通信网中使用分组交换也同样可以解决这个问题，参见6.3.1节。

4.1.4 利用MPEG压缩音频信号

MPEG的全称是动态影像专家组（Motion Pictures Experts Group），它是代表ISO和IEC对声音和图像压缩进行标准化的组织。目前已有不同的MPEG标准：

1）MPEG-1是为CD中播放音频和视频而制定的标准。

2）MPEG-2主要应用在音频和视频广播和DVD（Digital Versatile Disc）中。

3）MPEG-4涉及Internet和多媒体应用。

MPEG-3是为高清晰度电视（High Definition Television，HDTV）而制定的，但被集成到MPEG-2中。

音频信号至少从技术的角度来说比上节所介绍的话音信号的内容更丰富。它包括音调、噪声、音乐等频率范围为20kHz。这些信号用16Bit进行量化，采样频率为48kHz（如录音棚和数字音带）、44.1kHz（如CD）或32kHz（如数字卫星广播）。一个音道（单音道）的数据传输速率为512～768kBit/s。

不同的音频信号的特性区别很大，因此减小冗余带来的效果不是很好。于是转向研究信宿，即人耳，所以采用减少不相关的方法来实现。为此可以充分利用生理上的听觉效果。我们人类的听觉有下面的“不足”：

1）静音障碍和音响强度敏感都与频率有很大的关系。因此对不同的音频频谱范围要有不同的处理方法，无需传输听不到的声音。

2）将不同音调混合起来给人听，人耳并不能分辨出每个音调。这种所谓的阻塞效应与频率差别、音响差别和音调高低有关。通过对大量人员进行测试，研究者对这种效应进行了深入的研究。如果我们能识别阻塞效应，就可以指出与频率和声强有关的Mith，这样在发送器中就可以将听不到的声音去掉。图4-8所示为掩蔽阈的主要原理。

3）经过较强的耳压之后大约有10ms的时间，人耳几乎什么也听不到，然后人耳的听力逐渐恢复，大约在200ms之后才恢复到原有的敏感度，这种现象称为掩蔽。所

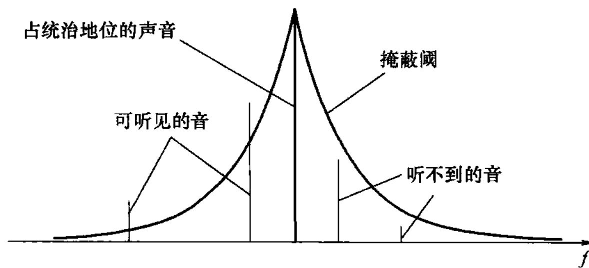


图 4-8 掩蔽阈的主要原理

以仔细地分析听觉事件肯定能降低数据传输速率。

所有这些影响在 MPEG 音频信号压缩中已得到了应用。图 4-9 所示为编码器的基本结构。

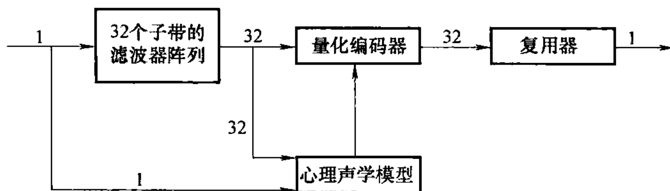


图 4-9 MPEG 音频编码器。所给数字为信号的个数

采样频率为 48kHz 的用 PCM 表示的音频信号在数字多级滤波器阵列中被分成 32 个子带，每个子带的带宽为 750Hz。根据心理声学模型编码器计算出每个子带的掩蔽阈，并确定每个子带的量化级别高度，以便保证量化噪声位于掩蔽阈处。根据音频信号的不同，量化器后的传输速率会有所变化。

32 个数据流在复用器中进行混合。这里包括采样值中发生波动的部分及表明实际字宽的常数部分。

如果需要存储被压缩的数据，那么就需要使用上面所描述的方法。因为绝大多数信息信道的带宽是通过最大数据传输速率定义的，当数据传输速率可变时，有时会降低带宽。基于该原因，在上面描述的处理过程中会将量化级别的数量提高，以便满足最大分配的数据传输速率。

附加的传输速率的提高并不能改善主观上的音频质量。还有一个原因，就是数字信号在处理的过程中受算术逻辑单元四舍五入等的影响也会变差。如果现在就将音频数据压缩到音频质量下降感觉还不很明显的程度，那么经过后续的处理，就会感到音频质量会下降得很明显。这种后续的处理包括音色调节、均衡以及解码后又进行的编码（级联编码、重编码）等。

因为解码器并不使用心理声学模型，它比编码器的实现简单得多。

上面所述就是编码的基本原理。这种方法称为 MUSICAM（Masking Pattern Universal Subband Integrated Coding And Multiplexing）。在 MPEG 标准中使用这种 MUSICAM 方法定义了所谓的 3 个层次。编码器和解码器的复杂性随层次的增加而提高，而传输速率却随之下降。对所有的 MPEG 方法来说，人们都特别注重让解码器比编码器更简单。

MPEG 方法目前已用在 DAB（Digital Audio Broadcasting）和 DVB（Digital Video Broadcasting）中。这种分布式业务需要大量的终端设备，因此设备的价格应尽可能低。

图 4-10 所示为 MPEG-2 第一层音频编码器的结构。

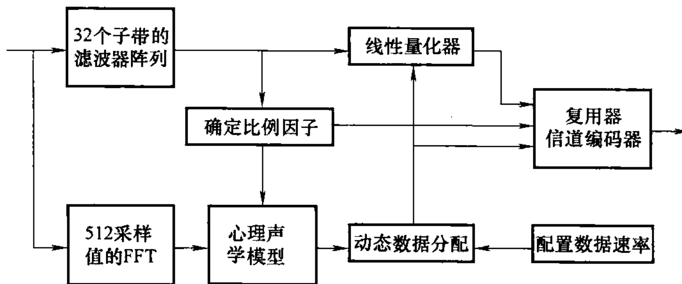


图 4-10　MPEG-2 第一层音频编码器（单声道）

图 4-10 中编码器输入端的采样频率值可为 48kHz、44.1kHz 或 32kHz，这里我们只关注第一种情况。

滤波器阵列之后确定了欠采样的 32 个输出信号（10 个为一组），之后的采样频率为 48kHz/32 = 1.5kHz。每 12 个相邻的采样值组成时长为 8ms 的一个块，每块的最大值即可确定。将每块的最大值作为该块的比例因子，并用 6Bit 进行量化。然后每块的所有 12 个值都与自己的比例因子相除，即进行归一化。

然后对各块归一化的采样值进行量化，将相应的量化级数的数量送给心理声学模型。所有的 12 个值都在同一时刻进行量化，由于掩蔽效应的作用这是可行的。因为掩蔽效应的持续时间最小是 10ms，大于每块的时长。

与此同时一个 FFT 实现对 512 个采样值的变化。它的频谱解析率比滤波器阵列之后的频谱解析率要高许多。在 FFT 的频谱中寻找等距离线，即确定存在的是音信号还是噪声。这个过程同 LPC 话音编码的信源选择过程是相同的。此时了解音频信息是必要的，这样心理声学模型才能够正确确定掩蔽阈的值。

这个过程之后是数据分配，即每个子带使用单独的量化级数，这样数据传输速率才能达到最大。每个子带有 4Bit 用来表示量化级数的数值，因此共有 16 种可能性。

最后复用器将所有的数据流复用在一起。图 4-11 所示为一种帧结构，它包含原始 PCM 信号 32 × 12 = 384 个采样值，时长为 8ms。

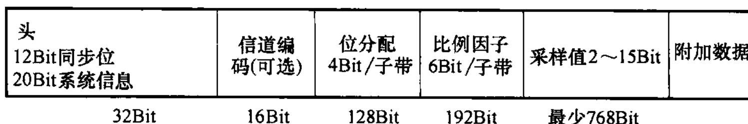


图 4-11　MPEG-2 第一层帧结构（单声道）

帧头由 12Bit 的同步位（参见 2.2 节的结尾部分）和 20Bit 的系统信息组成，它们给出音频信号的相关信息，如用于节目选择等。

可选择的差错保护部分使用的是卷积码，参见 4.3.3 节。这种信道编码保护的是位分配和部分头信息。

解码器并不使用附加数据部分。附加数据用于将来扩展，如给环绕立体声等使用。

只要这样的帧在误码率低于 10^{-3} 的信道中传输，人就不会感觉到干扰。

共有 14 种数据传输速率可供选择，它们的值为每音道 32～448kBit/s。

MPEG-2 第二层的压缩效果要好些，但编码器和解码器都比较复杂。数据传输速率是 32～384kBit/s。比较重要的变化是：

1）每个子块的长度用 36 个采样值代替 12 个。因此每秒钟需要的比例因子会降低。当音频信号变化较快时（击鼓声）每个子块的周期比掩蔽的时间要多很多。这种情况下每子块每个子带需要 2 个或 3 个比例因子。

2）FFT 用 1024 个采样值代替 512 个采样值，能提供较好的频率解析度。

3）在上述的子带上数据分配只有 4 种而不是 16 种可能性。

第三层编码又进了一步：

1）用 576 个子带代替 32 个。

2）采用非线性量化。

3）用霍夫曼编码减少冗余。

4）在立体声广播中信道之间的冗余也被消除掉。立体声不是简单的双声道声音，因为后者的两个声道相互之间完全没有关系。

4.1.5 利用 JPEG 压缩视频信号

我们先来看一下关于图像和视频的描述。

图像是二维信号。整幅图像被分成许多个像素（Pixels，picture element），每个像素有色度和亮度。根据色彩学的知识，每种颜色都是由 3 种基色组成，因此为了描述一幅图像需要 3 个矩阵 $\boldsymbol{R}$、$\boldsymbol{G}$、$\boldsymbol{B}$ 来分别代表红色、绿色和蓝色。

通常要对这些矩阵进行线性变换。首先构造一个和，生成带有亮度信息的矩阵 $\boldsymbol{Y}$：

$$\boldsymbol{Y} = 0.3\ \boldsymbol{R} + 0.59\ \boldsymbol{G} + 0.11\ \boldsymbol{B} \tag{4-7}$$

上式中的系数是根据人眼对各种颜色的敏感度来确定的。现在构造两个差：

$$\begin{aligned} \boldsymbol{U} &= \boldsymbol{B} - \boldsymbol{Y} \\ \boldsymbol{V} &= \boldsymbol{R} - \boldsymbol{Y} \end{aligned} \tag{4-8}$$

这样除了矩阵 ***RGB*** 之外又有了 3 个矩阵 ***YUV***。这种转换是一一对应的，因此两种描述方法是等价的。

用 ***YUV*** 来描述的优点有两个：

1）亮度信号 $\boldsymbol{Y}$ 使与黑白设备的兼容成为可能。

2）在 4.1.2 节中已经介绍了，人眼对颜色的敏感度不如对亮度的敏感度好，因此矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 能消除不相关。

二维信号不适合传输，因此人们将矩阵用按行取的办法将它转换成一维信号。在信源编码中，涉及到处理过程，这在矩阵描述中也能遇到。

视频由一系列的静止图像组成。每秒钟至少出现 25 幅图像给人的感觉才是连续的。

图像和视频是按照下面的标准进行编码的：

1）JPEG（Joint Photographic Experts Group）：静止图像。

2）M-JPEG（Motion-JPEG）：视频（活动图像），没有准确定义。

3）MPEG-1：CD-ROM 和多媒体使用的视频，数据速率可达 1.5MBit/s。

4）MPEG-2：电视和摄影棚使用的视频，数据速率为 5…50MBit/s。

5）MPEG-4：基于对象定义的互联网应用和多媒体技术。

在本节中我们只关注用来压缩静止图像的 JPEG 方法。这里会使用离散余弦变换，它是二维傅里叶变换的一种特殊形式。

最基本的想法我们用长度有限的一维信号来解释。例如一周内的温度信号，这个信号由一长串采样值组成，它的结构是事先定义好的，如按照昼夜这个规律。现在对这个序列进行离散傅里叶变换（DFT），可以得到频谱成分。它们中的两个值将占主要地位，即表示平均温度的 0Hz 成分和描述昼夜天气变化的日期数成分。其他所有成分都变得很小甚至可以忽略。

现在我们将这个例子用在图像信号上，它与时间无关但与地点有关，当然从数学的角度来说是无关紧要的。但是图像是二维的，所以必须使用二维傅里叶变换。对于离散信号使用来说 DFT（离散傅里叶变换）和 FFT（快速傅里叶变换）提供了很有效并且很快速的算法。

由于受边缘效应的影响，DFT 对周期信号来说会逐渐变弱。因此可将图像在两个方向上放大，可以采用添加镜像原始信号的办法。这样信号在原始域为偶数，没有跳变，并且可以周期性地延伸。此外 DFT 频谱是实数，这对减少计算的复杂度很有利。上面描述的对 DFT 的扩展也可以直接使用，也基于离散余弦变换。它与 DFT 是紧密相关的。对二维信号而言，有如下的公式：

$N \times M$ 矩阵的二维离散余弦变换为

$$S(X,Y) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} \alpha(X)\alpha(Y)s(x,y)\cos\left[\frac{(2x+1)X\pi}{2N}\right]\cos\left[\frac{(2y+1)Y\pi}{2M}\right] \tag{4-9}$$

$N \times M$ 矩阵的反向二维离散余弦变换为：

$$s(x,y) = \sum_{X=0}^{N-1}\sum_{Y=0}^{M-1} \alpha(X)\alpha(Y)S(X,Y)\cos\left[\frac{(2x+1)X\pi}{2N}\right]\cos\left[\frac{(2y+1)Y\pi}{2M}\right] \tag{4-10}$$

这里各符号的含义为：

$s(x,\ y)$ 为原始域内坐标为 x，y 的图像信号；

$S(X,\ Y)$ 为变换后频域内坐标为 X，Y 的图像信号；

N，M 为矩阵的行数和列数；

x，$X = 0 \cdots N-1$；

y，$Y = 0 \cdots M-1$；

α 为权重系数。

$$\alpha(r) = \begin{cases} \sqrt{\dfrac{1}{N}} & 当\quad r=0 \\ \sqrt{\dfrac{2}{N}} & 当\quad r\neq 0 \end{cases} \tag{4-11}$$

离散余弦变换对实数 $N \times M$ 矩阵进行一对一的变换，即变换的结果也能逆向转换成 $N \times M$ 矩阵。在变换的过程中没有信息丢失。正向和逆向变换使用相同的算术逻辑单元来实现。

当然 DCT 需要时间，并且它的复杂度随矩阵大小的增长而迅速增加。因此在图像压缩中将图像的大小拆成 8 像素乘 8 像素大小的子块，对每一子块使用 DCT。在 $N = M = 8$ 的特殊情况下式（4-9）可变为

$$S(X,Y) = \frac{1}{4}\beta(X)\beta(Y)\sum_{x=0}^{7}\sum_{y=0}^{7} s(x,y)\cos\left(\frac{(2x+1)X\pi}{16}\right)\cos\left(\frac{(2y+1)Y\pi}{16}\right) \tag{4-12}$$

其中的权重因子 β 满足：

$$\beta(r) = \begin{cases} \frac{1}{\sqrt{2}} & 当 \quad r = 0 \\ 1 & 当 \quad r \neq 0 \end{cases} \tag{4-13}$$

DCT 将一个子块变成一个基本结构，就像 DFT 将一个数字序列变成它的谐波成分一样。图 4-12 给出的是一个 4×4 矩阵的基本结构的例子。

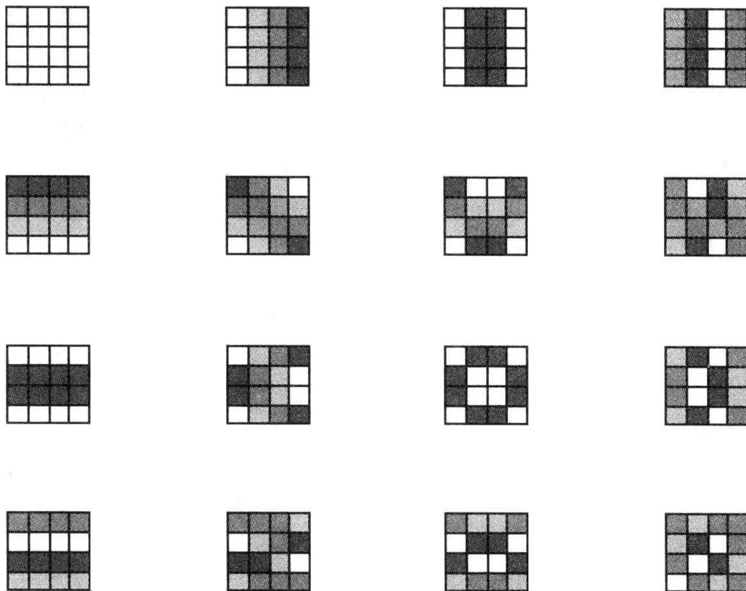


图 4-12　4×4 DCT 的基本函数

图 4-12 的解释如下：我们看到 16 个 4×4 矩阵，其中的黑色区域表示它所包含的数值。所有这些矩阵都在它的原始域内。其中的每个矩阵都属于一个 4×4 的频谱矩阵，它只有一个元素是 1 而其他的所有值均为 0。这唯一的一个 1 的位置由所给出的矩阵的位置来确定。

我们先来看最上面的一行，即一个一维信号。矩阵的最左端是一个频谱向量 [1 0 0 0]，这其中只出现一个直流成分。因此整个原始域矩阵的亮度相同。相邻的一个矩阵的频谱向量为 [0 1 0 0]，基波成分很活跃，因此原始图像区域有一个周期位置。再

向右，周期变短，图形变化加快。

最上面一行的所有矩阵在垂直方向上都没有固定结构。上数第二行最左面矩阵在垂直方向上有一个周期，就像最上面一行左数第二个矩阵在水平方向上有一个周期一样。

最右下角的矩阵毫无规律。

解释完 DCT 的机制之后，就会出现这样的问题，这样做有什么好处。为此我们看一个 8×8 子块如何在 JPEG 中应用的例子，这里仅给出所有的矩阵中的整数元素。

矩阵 1：原始图像域的子块

165	156	160	170	171	168	159	152
140	132	136	135	134	145	127	120
131	129	127	128	128	128	128	127
176	171	185	203	206	203	193	178
127	127	127	127	127	122	117	119
128	127	127	127	125	121	115	114
124	122	122	120	120	121	124	119
127	127	127	127	126	126	123	118

现在执行二维离散余弦变换：

矩阵 2：DCT 后频域内的子块：

1108	11	-22	14	2	5	4	-1
88	-4	-9	9	3	3	4	-2
-32	6	7	3	-3	1	1	-1
-18	10	1	-1	3	-2	0	0
95	-8	-15	2	5	2	-2	1
91	-16	-7	2	3	5	-4	2
-19	5	4	-5	3	-2	-3	0
-84	5	17	-7	0	-4	-3	2

这里引人注意的是在矩阵 2 中各个数值相差很大，而在矩阵 1 中各个数值几乎相等。经过 DCT 后这种矩阵特点是很典型的，这正好是引入减少不相关的切入点。

可能会有这样的想法，干脆将很小的值作为 0 处理。但这会引起较大的图像误差。经过许多人的大量实验，人们凭经验可以确定每个频谱成分的重要性。据此可以确定量化矩阵 $\boldsymbol{Q}$，用它来确定系数的权重。描述亮度的矩阵 $\boldsymbol{Y}$ 用 $\boldsymbol{Q}$ 来确定权重（归一化）然后进行四舍五入处理。这里使用的是针对矩阵中每个元素的除法，而不是整个矩阵的除法。

$$\boldsymbol{Y}_{\mathrm{Q}} = \mathrm{round}\left(\frac{\boldsymbol{Y}}{\boldsymbol{Q}}\right)$$

减少不相关就发生在四舍五入这一步。矩阵 3 是 JPEG 中提到的但没有任何规定的量化表。对颜色差别矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 使用另外一个表，它左上角区域的数值较小，而在其

他区域数值较大。因此高频成分的系数会减小，彩色图像的空间解析率会受到限制。实际使用的量化表必须与压缩的数据流一起传送给接收器，这样接收器才能正确解码。

矩阵3：量化表 $\boldsymbol{Q}$

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

矩阵4：经过加权和四舍五入的图像子块（频域范围）

69	1	-2	1	0	0	0	0
7	0	-1	0	0	0	0	0
-2	0	0	0	0	0	0	0
-1	1	0	0	0	0	0	0
5	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-1	0	0	0	0	0	0	0

矩阵4给出了与矩阵2相比量化的有效性。许多系数变成了0，这些0组合在与左上角较远的地方。

现在可以读取系数。按照图4-13所示的顺序，这样多个0在串行的数据流中可以放在一起。

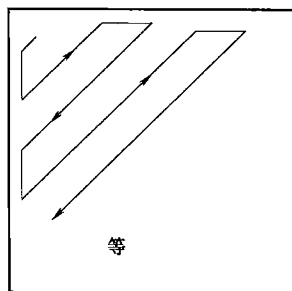


图4-13　从频谱矩阵中读取DCT系数

现在我们得到了由64个数字组成的表示一部分图像的串行数据流。当然还要对这些数据进行减少冗余：

1）因为当许多0连在一起时，通常显示地使用行程编码。

2）同时结合使用霍夫曼编码。

3）每个序列的第一个值表示整个子块的平均亮度，并且绝大多数情况下占有统治地位。为了减少这个值，对系数进行预编码。

然后对去掉不相关和冗余的数据流进行复用。

1）到现在为止我们只谈到了8×8的矩阵。一幅包含例如720×576像素的图像有6480个这样的8×8子块，每个子块需要3个矩阵，即 $\boldsymbol{Y}$、$\boldsymbol{U}$ 和 $\boldsymbol{V}$。

2）还需要一些关于使用的量化表格、图像的解析率等的详细信息。

数据流在接收端被解复用，并且恢复冗余。因此会重新得到矩阵4。该矩阵首先与 $\boldsymbol{Q}$ 进行复用（矩阵5），然后进行反向二维DCT运算，得到矩阵6。

矩阵 5：在解码器中对图像子块（频域）去归一化，比较矩阵 2

1104	11	-20	16	0	0	0	0
84	0	-14	0	0	0	0	0
-28	0	0	0	0	0	0	0
-14	17	0	0	0	0	0	0
90	0	0	0	0	0	0	0
96	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
-72	0	0	0	0	0	0	0

矩阵 6：反向 DCT 运算后的图像子块，比较矩阵 1

165	165	167	170	171	167	157	150
126	126	128	133	135	133	125	119
127	128	130	135	139	138	132	127
183	182	183	187	190	189	184	179
127	125	124	125	126	124	119	114
132	129	126	126	127	125	120	115
126	123	121	121	123	123	120	116
126	123	122	123	127	129	127	124

最后图 4-14 从视觉上给出了矩阵 1 和矩阵 6 的比较。

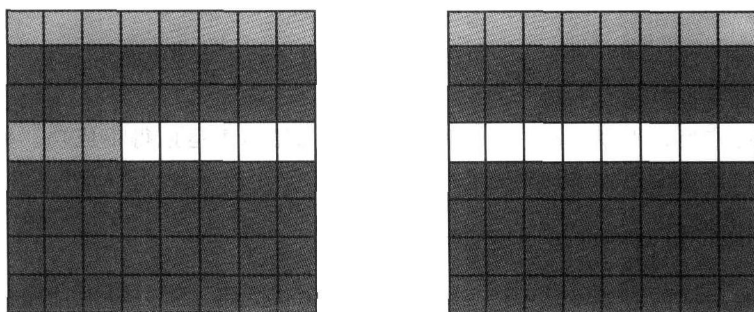


图 4-14 原始图像子块（左）和经过 JPEG 压缩后重新构造的图像子块（右）

在感觉不到图像质量变差的情况下使用 JPEG 压缩方法时，压缩系数可以达到 4～8。

当每个子块大于 8×8 像素时会比较好。但 DCT 的计算开销随矩阵的增大而超线性的增长，因此需要进行一些折中。

这里还值得提一下 JPEG 方法的一些特殊性：

1）若用互联网传输图像，人们还得接受恼人的等待时间。很显然传输能力要比目的地的计算能力小。为了适应这种环境，可以修改复用方法，这样无需对上面所描述的 8×8 子块的 DCT 系数进行传输（串行复用）。很多情况下可以先传所有子块的元素（1，1），即 DC 值，然后传输元素（2，1）、（2，1）等。这种过程复用（progress Mul-

tiplexing）的优点是解调器一开始就有整幅图像的所有信息，尽管还很不清晰。然后解码器不断地改善图像的质量，直到达到最优的质量。这对接收者来说比缓慢地从上到下一点一点地看到图像更有吸引力。此外当人们不希望继续观看这幅图像时，可以中断连接。类似地可以实现连续的近似，此时开始传输 DC 值，然后传输所有系数的 MSB，最后是不重要的数据信息。

2）可以使用量化矩阵来确定压缩系数以及与此相关的图像质量。

3）特殊的图像，如 X 光照片，只允许有较小的图像质量的降低。因此只需考虑是否可以减少冗余。理论上可以简单地将 $\boldsymbol{Q}$ 的所有元素都置成 1。但实际上会出现误差，这是由计算 DCT 及 IDCT 过程中的四舍五入引起的。因此 JPEG 允许使用有预编码的有损压缩。此时每个像素被它的 3 个邻居点保护，共有 8 中预编码方法：

C B
A X

Auswahlcode：	Pradiktion für X：
0	keine
1	A
2	B
3	C
4	A + B − C
5	A + (B − C)/2
6	B + (A − C)/2
7	(A + B)/2

编码器选择最好的一种预测，并将它通知给解码器。估计值还需额外使用霍夫曼编码。

使用这种方法的压缩系数在 2 以下。但 X 光照片的压缩系数能超过 2，因为照片中有很大一片都是黑色的。

离散余弦变换是很多变换中的一种，它在图像压缩中运用得很成功。另外一种比较新的、讨论得也很多的通过变换进行编码的方法是小波变换[Ban02][Str00]。

4.1.6 利用 MPEG 压缩视频信号

视频信号是活动的图像，它是三维的，即两个位置维度和一个时间维度。对人来说，每秒钟有 25 幅单个图像就足够了。每幅图像又按行分解，所有的行连在一起就构成了整幅图像。这样一个三维的信号就变成一个一维的信号。

彩色信号有多种呈现方式：

1）4: 4: 4 格式：用 13.5MHz 的采样频率分别对信号 R、G、B 进行采样，并使用 8Bit 的解析率。未压缩的数据传输速率为 R = 324MBit/s。

2）4: 2: 2 格式：信号 Y(f_A = 13.5MHz)，信号 U 和 V（分别为 6.75MHz），每个信号都使用 8Bit 的解析率，R = 216MBit/s。

3）4: 2: 0 格式：信号 Y(13.5MHz)，信号 U 和 V（分别为 3.375MHz），每个信号都使用 8Bit 的解析率，R = 162MBit/s。

从 RGB 描述转换到 YUV 描述已经引起减少不相关，即对颜色的限制。4:4:4 格式

对图像的处理是比较有利的，例如为了产生魔术效果。

下面3个原因使JPEG不适合视频压缩：

1）所谓的运动JPEG方法（M-JPEG）没有被标准化，因此会出现兼容性问题。

2）JPEG不使用视频图像时间上的相关性。

3）JPEG还不能预见到音频和视频的联合编码和复用。

MPEG没有上面的限制。正如在4.1.4节中所介绍的那样，MPEG有不同的标准：

1）MPEG-1：应用领域是计算机上的多媒体应用，因此会出现不同的限制。数据传输速率为1.5MBit/s，因此很适合传统的CD。图像格式最大为352×288像素。

微电子技术的进步很快地使标准过时，因此出现了像MPEG-1+或MPEG-1.5这样的名字。

2）MPEG-2：从最开始这个标准就是为电视传输而设计的。它提供不同的质量级别，称为Profiles（类）和levels（级别）。Pal格式的电视信号（参见5.1.2节）可使用6MBit/s的速率（终端用户质量），9MBit/s的速率（摄影棚质量）为进一步的处理留有余量。

MPEG-3应该规定高清电视（High Definition TV，HDTV）的编码标准，目前这些内容都被包含在MPEG-2中。

3）MPEG-4：为基于互联网的多媒体应用而设计的。它所必需的压缩编码方法采用了一种新的技术，使用对象来描述多媒体的内容。

MPEG-1、MPEG-2同JPEG一样都是基于DCT的，但它们需要进一步的处理：

1）差分编码：相邻的图像采用差分编码，它的原理在图3-47中已经描述过了。因为每行的所有点以及图像的所有行已经使用DCT进行了去相关，所以预测器的延迟包含整幅图像的信息。

2）运动估计：预测中还包含对运动的估计。它包含16×16像素的宏块，即2×2个DCT的子块。运动估计要确定哪些宏块之间是相似的，以便在这些块之间可以使用差分编码。图4-15给出了MPEG视频编码器的结构图，这里运动估计的影响放在了预测器中。运动估计需要大量的计算，但这些计算只集中在编码器中，对解码器复杂性的增加是比较小的。

3）图像的重新分类：预测器不仅仅使用过去的图像，也使用将来的图像（双向预测）。该过程会产生一定的延迟，并要求编码器和解码器有较大的内存。因为预测编码会预埋一些误差，并且电视接收器有可能在节目的中间开机，所以偶尔会传输没经过预测的整幅图像，即所谓的I帧（帧内编码图像）。在I帧之间放置P帧（单向预测帧）作为支持。绝大多数的帧是B帧，即双向预测帧。帧的顺序如下：

$$\cdots I_k \quad B_{01} \quad B_{02} \quad P_1 \quad B_{11} \quad B_{12} \quad P_2 \quad B_{21} \quad B_{22} \quad P_3 \quad B_{31} \quad B_{32} \quad I_{k+1}\cdots$$

即共有12幅图像，到新开机的接收器能够开始解码的持续时间约为半秒，这半秒对最快的遥控器也足够了。

为了在解码器中能重新构造出B帧，需要使用相邻的I帧和P帧。因此在编码器中对帧的顺序进行调整：

$$\cdots I_k \quad P_1 \quad B_{01} \quad B_{02} \quad P_2 \quad B_{11} \quad B_{12} \quad P_3 \quad B_{21} \quad B_{22} \quad I_{k+1} \quad B_{31} \quad B_{32}\cdots$$

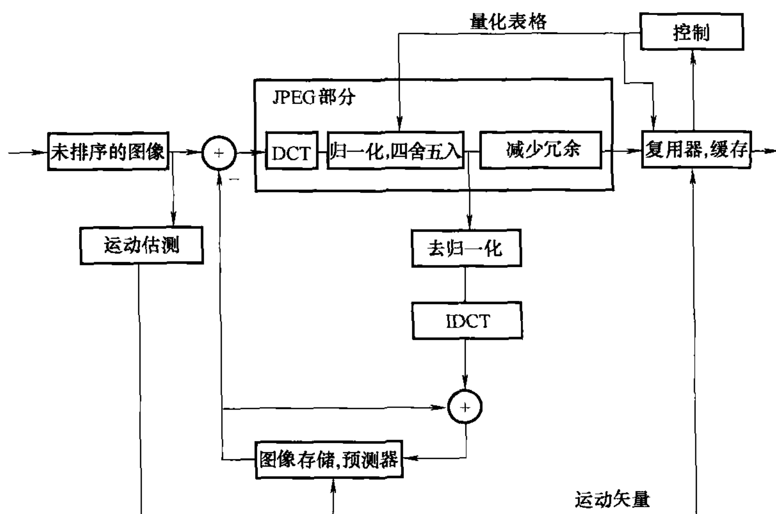


图 4-15 MPEG 视频编码器的结构图

这种方法的优点在于在解码器中只需要存储 2 帧而不需要存储 4 帧，因此解码器可以便宜些。

4）用可变量化来控制数据传输速率：它的目标是充分利用当前传输信道所能够提供的数据传输速率。压缩深度与视频信号有关，因此量化矩阵 Q 动态地对图像内容进行适配。当然当前使用的 Q 也必需通知给解码器。同 JPEG 不同，在 MPEG 中预先定义了几个版本的 Q，所以只需传输所选版本的号码即可。场景变化时会看到很短的人为效果，但是在一个场景内观察不到任何编码的影响。

MPEG-1 使用的格式被称为 SIF（Source Input Format），它的数据传输速率只有 4:2:0 格式的一半，这是以出现误差为代价的。

MPEG-2 使用的格式是 4:2:0 和 4:2:2。除此之外还考虑了跳行方法（参见 5.1.2 节）并且可以使用非线性量化作为替代。表 4-3 给出了 MPEG-2 视频编码的类和级别，其中级别给出了图像的解析率，类包括：

1）Simple Profile（简单类，SP）：只有单向预测，即没有 B 帧。格式为 4:2:0。

2）Main Profile（主类，MP）：同 SP 类似，但额外附件双向预测。MP 是为绝大多数常见的应用而设计的，对数字电视（DVB）也适用。

3）SNR Scalable Profile（信噪比可分级，SNRP）：数字传输从完美的传输到完全瘫痪只有很小的差别，如图 4-16 所示。在过渡区域 MPEG-2 可以选择，即希望有点噪声但很清晰的图像还是希望没有噪声但不清晰的图像。SNRP 情况下可以获得第一种情况。

4）Spatial Scalable Profile（空间可分级，SSP）：在过渡区域得到的是不清晰的，但没有噪声的图像。

5）High Profile（高类，HP）：最高的需求采用 4:2:2 格式。

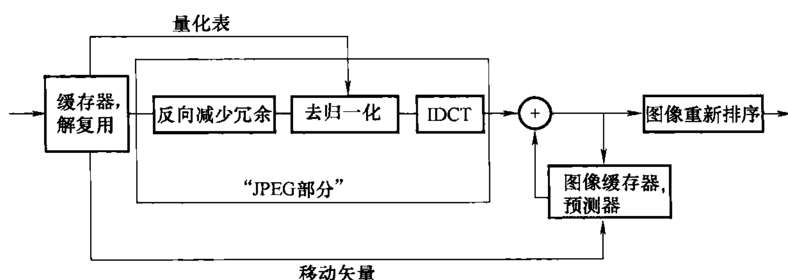


图 4-16　MPEG 视频解码器的结构图

表 4-3　MPEG-2 视频编码类和级别的最大速率

类 级别/解析率	Simple Profile	Main Profile	SNR Scalable Profile	Spatial Scalable Profile	High Profile
High Level 1920 × 1152		80Mbit/s			100MBit/s
High-1440 Level 1440 × 1152		60MBit/s		60MBit/s	80MBit/s
Main Level 720 × 576	15MBit/s	15MBit/s	15MBit/s		20MBit/s
Low Level 352 × 288		4MBit/s	4MBit/s		

MPEG-2 能满足电视传输的所有需求：

1）Low definition Television（低清晰度电视，LDTV）：类：SP@ML，传输速率：1.5～3Mbit/s。

2）Standard definition Television（标准清晰度电视，SDTV）：MP@ML，传输速率：3～6MBit/s。

3）Extended definition television（EDTV，扩展清晰度电视）：HP@ML，传输速率：6～8MBit/s。

4）High definition television（HDTV，高清晰度电视）：HP@ML，传输速率：20～30MBit/s。

数字电视（DVB）计划使用 MP@ML。

表 4-3 中空着的域还没确定下来。该表格是向下兼容的，即某一个域的解码器对它左面和下面的域确定的模式也同样能解码。

MPEG-2 也定义了同属视频压缩数据和音频压缩数据的复用。这需要缓存和同步机制，这样解码后信号可以同时提供给信宿。除了直接的时间上的关系外，这样做还要求其他信息，如节目的类型和字幕等。复用后的信号称为 transport stream（传输流），如图 4-17 所示。

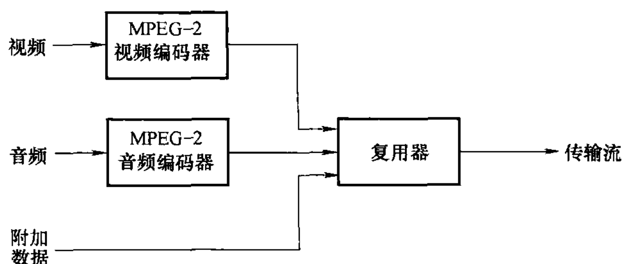


图 4-17 MPEG-2 中视频、音频和其他数据的复用

4.2 加密

4.2.1 引言

加密将明文（plaintext）转换成密文（ciphertext)。这样做的目的有多个：

1）私密性：避免未被授权的人阅读信息。

2）完整性：避免修改或删除未被授权的信息。

3）认证：验证信息确实是从希望的发送者发出的并且去往希望的目的地。

加密的第一个应用来自军事和外交上的需求。目前私有企业，如银行对私密性也很感兴趣。以前使用表格或机械设备进行加密和解密，现在使用计算机。加密器就是一段程序。

每个加密过程都使用一个算法，即一个特殊类型的加密器，和一个密钥，它用来设置加密器的参数。目前的方法仅依靠密钥就能保证安全性。危险人员（密码学中称之为未被授权的第三者）允许使用合理的解密器，并且也允许他了解加密算法，尽管如此没有密钥他仍不能获得明文。这种方法的优点是只需开发一个加密算法即可。

密码学应用中的一个难点就是如何安全可靠地分配密钥。

密码分析学是研究如何获取密钥的一门艺术。从原理上说对每个密文，通过将密钥的所有可能性都试一遍来获取密钥是可能的。但由于众多可能性的存在这种方法在实际中是不可行的。人们已经使用了一种方法，地球上所有的计算机同时计算几千年才能安全地获取明文。密码分析家于是寻找一种耗时较短的方法以便能获取密钥。因此密码学使用富有挑战性的数学方法，是一个有少数专家的领域。现代系统密钥是自动产生的并且没有人知道。

人们进行如下的区别：

1）在线加密：在发送前直接进行加密，接收后直接解密。这种方法很简单，但不灵活。

2）离线加密：加密和传输在时间和地点上是分离的，转接人员既不知道密钥，也不了解算法，他只负责处理密文。

一段文字会使用二进制码来描述，例如大家都很熟悉的 ASCII 码。加密并不对字

符进行处理，而是对数据的组合进行处理。使用同样的方法也可以对其他的信息，如数字话音或图像进行加密。

4.2.2 对称加密方法

对称加密方法指的是加密和解密使用两个相同的密钥。非对称加密使用两个不同的密钥，我们将在 4.2.3 节中介绍。

对称加密方法分成两类：替代（Substitution）方法、置换（Permutation）或变换方法（打乱顺序）。

在替代算法中 26 个字母中的每一个都用其他的字母来代替，如 a 用 f，b 用 x 来代替等。替代字母的分配表由密钥来确定，即共有 $26! \approx 4\times10^{26}$种可能性。如果用计算机来探测所有的可能性，每种可能性需要 1μs，那么最多需要 10^{13}年，平均需要这一半的时间。

这种方法看起来很安全，但却很容易被攻击。将密文的符号进行排序即可以得到每个字母出现的频率分布。德语或英语的明文中出现最多的是字母 e 等。也可以进一步对两个字母的组合（二连字，如“en”或“er”出现很频繁）和三字母组合，如“sch”或“die”进行分析。

将字母和字母组合的分布进行平衡的一种可能性叫做多字母加密。原则上在加密的过程中要对字母分配表不断地进行更换。

置换或变换方法指的是在加密的过程中字母的顺序会发生变化。例如文字可能几行作为一个矩阵来读入，几列一起来读出。字母出现频率的分布当然也会保留。置换方法改变位置时不改变符号的值；当替换值时，不修改符号的位置。

对符号（字母）的分布的研究出于类比的原因被称为频谱分析。因此加密的一个任务是在密文中能够保证各字母的等分布。这相当于一个白噪声，加密过程就像减少冗余一样是去相关。这带来一个优点，即密文与一个真的随机过程不能区分开来。如果使用明信道传输，窃听者不能从真实的数据中得到明文信息。这点很重要，因为在军事应用中找到密钥所消耗的时间远比信息本身的实时性长很多。因此人们经常从开始就消除解码的念头，而将精力集中在对流量的统计分析上。使用明信道传输会抵消它的作用。

目前所应用的是进一步发展起来的方法以及上面所描述的置换和替代方法的结合。最著名的例子是 IBM 在 1974 年开发的并被美国政府引为标准的 DES（Data Encryption Standard）算法。由于相关的市场很大，DES 算法也可用纯硬件来实现，允许较高的数据传输速率。例如 1994 年 1GBit/s 的 DES 加密可用 GaAs-Gate-Array（GaAs 门阵列）来实现。

DES 算法对每个 64Bit 组成的块进行加密，每经过 16 次迭代进行一次替换，同时还补充进行置换。密钥的长度为 56Bit，可以提供 $2^{56} \approx 7\times10^{16}$种不同的可能性。但这个密钥长度正是 DES 方法的问题所在。最新的研究表明，密码分析家使用大约 100 万美元的投资，可以在几个小时内破译 DES 密文。

在这点上三重 DES 方法较好，通过使用两个密钥，可以轮流执行 3 次 DES 算法。

苏黎世的 ETH 大学在 20 世纪 80 年代末开发出了 IDEA（International Data Encryption Algorithm）。这种算法展现出很大的声望。这是因为一方面他们的精神领袖 Lai 和 Massey 有公认的才能，另一方面瑞士不是 NATO 成员国，不受 NSA（National Security Agency，USA）的影响。IDEA 由不同的代数运算组合而成，它们可以由软件很有效地来实现。密钥的长度是 128Bit。

块加密的新标准称为 AES（Advanced Encryption Standard）。

作为上面所描述的块加密方法的一个变种，这种方法引入连续加密的方法（stream encryption）。此时二进制明文与密钥进行异或运算：

明文：	1	1	0	1	0	0	1	1	0
密钥：	0	1	1	1	0	0	1	0	0
密文：	1	0	1	0	0	0	0	1	0

在接收端使用相同的密钥再进行一次异或运算：

接收的密文：	1	0	1	0	0	0	0	1	0
密钥：	0	1	1	1	0	0	1	0	0
重新构成的明文：	1	1	0	1	0	0	1	1	0

只要密钥序列的长度至少与明文相等，并且只使用少数的几次，那么这种方法很简单并且很安全。该方法因此也被称为实时填充（on time pad）。

在所有的对称加密方法中，密钥必需要通过很安全的途径交给接收者，例如通过一个信使。这要在一个合适的时间点上进行，而实际信息的传输时间经常受许多外界因素的影响。密钥由实时随机序列产生，例如可以利用电阻的噪声等。

一种简单的交换密钥的方法过程如下：

A 用只有他自己知道的密钥对信息加密，并将它发给 B。B 同样用只有他自己知道的密钥对收到的密文再次加密并回送给 A。A 用他的密钥进行解密并将仍为密文的信息再次发送给 B。B 用自己的密钥对信息解码从而得到明文。但这种方法有两个问题。一是合作双方必须是可以识别的，否则一个不认识的私自闯进来的人可能会接替 B 的角色；二是这种方法多次占用往返的信道。

4.2.3 非对称加密方法（公共密钥系统）

公共密钥系统最初是由 Diffe 和 Hellmann 在 1976 年提出来的。他们用两个不同的密钥，即密钥 C 用来加密，密钥 D 用来解密。密钥的产生方法是 D 不能根据 C 推导出来。因此 C 可以公开出来，正如这种方法的名字所给出的一样。

如果 A 想向 B 发送秘密信息，那么它使用 B 的公开密钥 C_B 对信息加密，并且将密文发给 B。B 用他私有的密钥 D_B 进行解密。

这里最大的优点是无需交换私钥 D。这允许自发保护的连接接受一种新的保护方式。此外必要的密钥数量会减少，因为每个用户只需要一对密钥就够了。在对称方法中每个连接就需要一个密钥。

当然公共密钥系统也有缺点：

1）对较长的信息来说加密的复杂性很高，不实用。因此公共密钥系统主要用来交

换对称加密方法的密钥。

2）公开密钥的发布必须与认证相结合。如果一个敌手伪造了 B 的公钥，那么他可以容易地将发给 B 的信息进行解密。

3）非对称方法的安全性基于像对一个大数进行因式分解那样的数学难题。它是否这样难还不能证明。如果已经发现了一种有效的求解算法，那么为了安全起见，需要尽可能长时间地对它保密。因此对加密来说发现不断增大的质数也是很有意义的。

最有名的非对称加密方法就是由 R. Rivest 和 A. Shamir 公布的 RSA 算法。

4.2.4 哈希函数

哈希函数是具有下列特点的单向函数：

1）从任意长度的信息中可以产生一个长度固定的并且较短的哈希值。

2）哈希函数没有冲突，即使信息的内容相差很小也能给出不同的哈希值。

3）哈希值的计算很简单很快。但在实际中不可能从一个事先给定的哈希值产生一段信息。

使用哈希函数可以对信息进行压缩并将它变得不可识别，但却不能重新对它扩展变得可读。一种使用很广但同时却不是很安全的算法是 MD5 算法。更现代更安全的是 SHA1（Secure High Algorithm）。

哈希函数被用来对信息进行安全认证，方法是将经过加密的哈希值作为校验和或指纹附在信息上。以前经常做的是给信息加上一个时间戳（time stamp），以计算时延。在这种方法中，信息对所有人来说都是可读的，但它可以防止例如计算机程序对信息的内容进行修改。对整个信息进行加密当然也可以实现认证，但这显然比只加密较短的哈希值要复杂得多。

另一种应用是数字签名，参见 4.2.5 节。

计算机系统和芯片卡使用哈希函数来对密码进行保护。密码本身在系统中或卡中是根本不存在的。未被授权的人最多只能看到哈希值，但却不能从中推导出密码。

4.2.5 密码协议

私密信息交换

私密信息交换的原理我们已经提到过了：A 使用如 DES 或 IDEA 的对称加密方法，为此他产生一个新的只使用一次的密钥，这个密钥使用示 RSA 等的公共密钥系统用 B 的公钥进行加密，然后将加密的信息和加密的密钥发送给 B。

由 Zimmermann 开发的在互联网上变得著名的 PGP（Pretty Good Privacy）算法就使用上面的技术，它将 IDEA 和 RSA 结合起来。

问题是如何对用户进行认证。一个敌手可能会假称是 B。一种办法是采用经过鉴定的密钥管理机构（电子公证）对公钥进行管理。它使用每个密钥的哈希值，并且使用他自己的私钥即可。这样通过管理机构中所拥有的密钥就能够证明，这个公钥是否是 B 的。

数字签名

加密也提供了通过电子途径进行公证的可能性，即使用电子签名对文件进行安全

的存储和管理。经过签名的文件有下面的安全性保证：

1）不论是发送者还是接收者都不能对它进行额外的修改。

2）发送者和接收者都不能否认他的存在。

3）签名者是可以唯一识别的，只有他才可以产生签名。

4）签名唯一的属于一个文件，不能在其他的地方被复制和传输。

对数字签名使用非对称加密方法进行加密。这点充分利用 RSA 算法中两个密钥的角色，可以进行交换。签名是这样产生的，文件或仅是它的一个哈希值使用发送者的私钥进行加密（相反保密信息的传输使用接收者的公钥）。通过使用发送者的公钥的哈希值对签名进行解密，并与 B 自己计算的哈希值对比，就可以对签名进行验证。

电子货币

被签名的文件可以是张支票。它可以被整体复制，目的是多次进行使用。在支票上要有时间戳，这样可以对多次使用进行记录。

货币还有另外一个特点，就是它本身也要通过加密的方法来实现，即货币是匿名的。若使用现金付款，买者本身是不能被识别的。使用 Chaum 在 1985 提出的协议，每个人都可以对电子钞票的真伪进行辨别，尽管银行从未发行过这种钞票。但银行可以验证，这张钞票是否被多次使用。

电子水印（*Steganogramme*）

水印是指不引人注意地传输信息。为此将信息进行加密并且嵌入在一个较大的文件中，如图形或声音文件。不论从光学上还是声学上来说这种改变都是不可察觉的。这样可以提供例如版权声明一样的功能。

关于加密的详细信息可以参考一些专业文献，如参考文献【Sch96】。作为兴趣阅读可以参见参考文献【Beu94】，它没有复杂的数学推导。在参考文献【Sta97】中也能得到进一步的启发。此外参考文献【Sin00】也是一本很令人兴奋的床边读物。

最后我们以一个数值例子来看一下加密的开销，这个例子基于下面的假设：需要对 2MBit/s 的数据线路或无线链路进行加密。两台加密设备合起来花费 20000 欧元。这些投资应该在 5 年内收回，即每个月加密的开销为 333 欧元。

这样的链路每个月共可传输数据 5200GBit。假设我们用相对较差的平均利用率 10% 来计算，那么传输的数据量为 520GBit，相当于 65GB。这又相当于 16000000 张写满字的 A4 纸。

如果使用这样的链路传输 100 张 A4 纸的文件，那么对这个文件进行加密的开销总共为 0.002 欧元。

4.3 信道编码

4.3.1 引言

信道编码的目的是保证数据的可靠性，即在传输和存储中对信息进行干扰保护。为此需要增加针对原始信息的冗余，这样接收者根据双份的信息至少能发现差错，甚

至能纠正差错。

1kBit 大小的信息通过信道编码之后，它的大小会增加，例增加到 1.3kBit。但在该过程中以 bit 为单位的信息量不会增加，因为信道编码只是增加冗余。但传输这 1.3kBit 的数据要么需要额外的时间（在脉冲宽度或符号速率不改变的情况下），要么需要额外的带宽（以系数 1.3 将符号缩短）。这种方法的优点是允许信道的动态性变小，误码率增加。信道编码与调制相同，都能对信息立方体（Nachrichtenquaders）进行变形，如图 1-35 所示。因此信道编码同样属于信道适配。

图 4-18 中下面的信道是信道编码器和信道解码器之间的信道。除了传输媒体外，该信道还包括调制器和放大器等。可用下面方法描述该信道：

1）数据传输速率 R，单位是 Bit/s；

2）误码率（Bit Error Rate，BER）；

3）差错的类型（单个差错，突发差错等）。

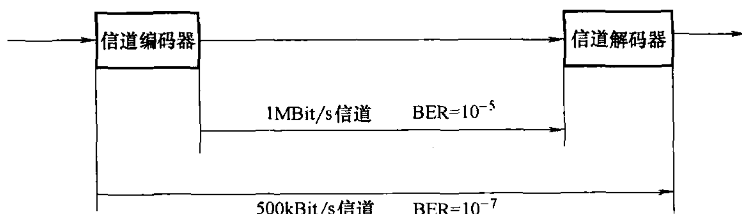


图 4-18 通过信道编码进行信道转换及信道适配

信源必须保证的可靠性即数据的误差是由应用来规定的。在 64kBit/s 的 PCM 话音传输中 BER 为 10^{-5}，平均每 1.56s 会出现一个误码。但这基本上不会产生任何影响，参见表 3-5。但一个 100KB（800kBit）长的计算机程序经过传输后从统计学的角度来看会出现 8 个 Bit 的错误，这文件几乎不能使用。

可能的误差源是多样的，它不仅仅包括机器设备，也包括人，例如在获取数据时。如果一个文本本身已经包含了许多输入错误，再使用 BER 为 10^{-9} 的信道来传输这个文本文件是毫无意义的。通常这些输入会导致文本的 BER 为 10^{-4} ~ 10^{-5}。

300 页的一本书大概包含 800000 字符。如果用 Baudot 码（每个字母用 5Bit）来表示这本书，并且用图 2-1 介绍的使用一个开始 Bit，两个停止 Bit 来传输，那么产生的数据量是 6.4MBit。当 BER 为 10^{-5} 时，大概有 60 个 Bit 会出现错误，即有 60 个错误的字母。因为在文本中会有一些冗余，所以这些错误很少会引人注意。

一个系统的差错概率是系统中每个单元的差错概率的和。对实际中经常使用的误码率，实际上人们已经忽略掉发生错误的数据后，又因为出错而变正确的概率。目前误码率已经能够达到 10^{-14}，这相当于在 64 kBit/s 的传输中，50 年会遇到一个错误。

最大的差错来源之一是数据收集，特别是当有人工操作时。因此在银行核对账单时，主要是对每个数据所填入的表格的位置进行控制，而不是检查总数是否正确。对人工输入数据进行测试的结果表明，大概会有 1% 的字符误差。进行两次输入，并对结果进行自动比较，会将 BER 提高到 10^{-5}。比较有帮助的是进行格式检查及产生校验和

等。第二种方法不外乎就是信道编码，而第一种方法是一个可信性测试。

表 4-4 给出了不同线路类型的误码率。这些值在不远的将来还会提高。

表 4-4 不同线路的 BER

线路类型	BER
拨号电话线路	$10^{-4} \sim 10^{-6}$
租用电话线路	$10^{-5} \sim 10^{-6}$
租用数据线路	$10^{-6} \sim 10^{-7}$

误码通常不会单个出现，而是一组出现（批误码或突发误码）。当用光纤传输数据时，突发误码会比用无线传输产生的突发误码要短。最长的突发误码是当接收端的解码器失去同步时发生的。存储设备（如 CD）也会产生突发误码，因为光盘表面的损伤或玷污会涉及到大量数据。表 4-5 给出的是对租用线路误码的一个评估。

表 4-5 测量到的一条租用线路的突发误码分布

受到干扰的相邻数据的数量	百分比
1	62.5%
2	22%
3	7.5%
4	4.6%
5	1.7%
6	0.9%
7	0.5%
≥8	0.3%

下面描述的是几个简单的保证数据可靠性的方法，它们还没涉及到复杂的数学问题：

1）奇偶校验：这是一个保证字符可靠性的方法，每个数据字都附加一个 Bit。这些奇偶位包含数据字中所有 1 的和（包括奇偶位）的极性，即是奇数还是偶数。这样接收器能够识别单个差错。在出现差错的情况下可以通过反向信道要求重传（Automatic Repeat Request，ARQ）。在差分编码中不能使用奇偶校验，因为会出现双倍的误码。

2）带有奇偶位的分组可靠性：将数据写入一个矩阵，每行和每列都添加一个奇偶位。因此这种方法也称为 VRC/LRC（Vertical Redundancy Check 及 Longitudinal Redundancy Check）编码。与按字符的奇偶可靠性保证不同，这种方法不仅能识别单个的错误，甚至还能纠正错误。但只能发现两个错误。这种好处是用较高的冗余来换取的。例如，4 个数据比特 $d_1 \cdots d_4$ 是通过 4 个验证比特 $p_1 \cdots p_4$ 来保护的。

$$\begin{matrix} \begin{bmatrix} d_2 & d_2 \\ d_3 & d_4 \end{bmatrix} & \begin{matrix} p_1 \\ p_2 \end{matrix} \rightarrow [d_1 d_2 d_3 d_4 p_1 p_2 p_3 p_4] \\ \begin{matrix} p_3 & p_4 \end{matrix} & \end{matrix}$$

当极性为偶数时，$p_1 = d_1 \oplus d_2$，$p_2 = d_3 \oplus d_4$ 等。接收者根据接收到的有用比特做相同的运算，获得4个验证比特 $p_{1e} \cdots p_{4e}$。然后将这些验证比特与收到的验证比特相比较。如果 $p_{1e} \neq p_1$，并且 $p_{4e} \neq p_4$，那么 d_2 是错误的。因为这些比特只能有两个可能值，所以直接将它翻转就能进行纠错。如果结果只是 $p_{1e} \neq p_1$，那么只是在传输的过程中将验证比特 p_1 进行了翻转，因而无需采取任何措施。

3）等权重码：一个码字的权重是它的位图中1的个数。等权重码只使用具有相同权重的字。传输中的单个错误会改变权重，因而会被发现。在实际中 CCITT 字母表 No. 2（电传打字机中使用的5位 Baudot 码）就延长了2位，因此在每个码字中如果出现了3个1，那么第四位就是0。CCITT 字母表 No. 3 使用新的码字。解码器按照3:4的比例进行检查，并且用 ARQ 信道进行确认。最有效的办法是3个字母构成块传输并且一起进行确认。这种方法主要用在使用短波无线信道传输电报的情况，因此也被称作 TOR（Teletype Over Radio）。在深海航行中使用的变种称为 SITOR（Simplex-TOR），业余无线电爱好者使用在实际中是相同的 AMTOR（Amateur-TOR）。目前还有许多新开发的变种正在使用。

4）回音法：接收者将整个消息反馈给发送者，列出其发现的可能的区别，必要的情况下可要求重新发送。使用这种方法无需传输冗余，但要求反馈信道必须与前向信道的能力相当。这种方法在终端中实现。特别是在传输数据较少的应用中这种方法是很有用的，例如从主计算机中下载一些参数到外部计算机中。此方法不要求接收端是智能的。

所有的 ARQ 方法都由协议来控制具体的运行场景。例如如果证实消息丢失，接收者也必须清楚，现在他收到的是对前面消息的重复还是一个新的消息。

上面描述的方法很容易理解并且很容易实现。但它们的作用是很有限的。有效的纠错方法使用复杂的数学方法来实现，例如用 CRC（Cyclic Redundancy Check）来保证可靠性。第4.3.2节和4.3.3节将具体介绍这些方法。表4-6给出了数据传输中由信道编码而得到的改善程度。

表4-6　有可靠性措施与无可靠性措施相比的改善程度

可靠性保证方法	改善因数	可靠性保证方法	改善因数
字符奇偶性	约100	CRC 可靠性	50000到100000
块奇偶性	约1000		

如果没有反向信道，就不能实现 ARQ 系统。这种情况也涉及到存储的情况，因为绝大多数情况下错误比特都是由存储引起的而不是由阅读引起的。因此重读一遍是毫无意义的。在这种情况下需要对信息增加足够的冗余，接收者不用反问就能对差错进行纠正。这种方法被称为 FEC（Forward Error Correction）。

信道编码从狭义上来说描述的只是数学上比较复杂的方法，它们可以分成分组码（4.3.2节）及卷积码（4.3.3节）两大类。前者起源于20世纪50年代，由于它较复杂，所以其建议的方法通常认为是不可用的。今天更复杂的方法至少可以用在 VLSI 电路中。最突出的例子还是 CD 播放器，其他例子还包括 HF、VHF、UHF 频段的无线数

据广播、无线移动通信（GSM 手机）、航天飞行以及遥测等。信道编码取得第一个巨大成功的例子是在航天飞船上的应用，图像从海王星上经过大约 40 亿 km 的距离返回地球。没有信道编码接收到的信号根本是不可用的。

4.3.2 分组码

分组码根据 m 个有用比特计算出 p 个检验位并将它们附在有用比特之后，这样新的数据字长度为 $n=m+k$，如图 4-19 所示。



图 4-19 长度为 $n=8$ 的码字，其中包括 $m=5$ 个有用比特，$k=3$ 个校验比特

m 和 k 的长度是任选的，并且不会根据被保护的数据的结构来确定。通常人们只是将有用数据当成简单的位图来看待。

长度为 m 的数据字可有 2^m 个不同的位图。图 4-19 所示的码字的长度虽然为 $n=m+k$，但它也只能产生 2^m 个不同的位图而不是 2^{m+k}个，这是因为 k 个校验位是采用固定的计算规则从 m 个有用位中产生出来的。

信道编码的基本思想是长度为 n 的码字共有 2^n 种可能性，但只能利用其中的 2^m 种。如果在传输中发生 1Bit 的差错，那么就会产生一个新的不允许出现的码字。因此就能发现这个差错。

一个很简单的例子就能将这个问题看得很明白。假设 $m=2$，$k=1$，那么共有 4 种不同的有用字组合，加上控制比特后会产生下面的可能的码字：

$$0\ 0 \rightarrow 0\ 0\ 0$$
$$0\ 1 \rightarrow 0\ 1\ 1$$
$$1\ 0 \rightarrow 1\ 0\ 1$$
$$1\ 1 \rightarrow 1\ 1\ 0$$

这里使用的是大家都比较熟悉的奇偶校验。通过实验可以确定，单个比特的误码绝不会将原码变成一个新的允许的码字。特别值得注意的是，如果将一个码字作为 n 维向量来观察，那么每个比特就是这个 n 维向量空间的一个点的一个元素。图 4-20 所示为 $n=3$ 的情况。在该图中对有效点进行了加黑。从这个图中可以看出有效码字总是在对角线上，绝不会在立方体相邻的两点上。如果一个码字出现了一个比特的误码，例如从 011 变成了 001，那么结果就是一个无效的码字。但这样做却不可能对这个误码进行纠正，例如最初发出的码字也可能是 101，出现单比特的误码它也可能变成 001。

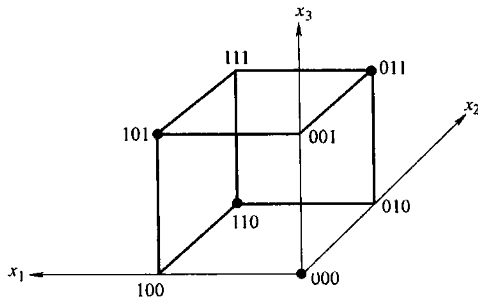


图 4-20 用 n 维空间来描述 $n=3$ 时的码字

图 4-20 也表明，使用到目前为止无效的码字可以构成另外一个码字，这两个码字是等效的。

这样码字产生的问题变成如下：从$2^n=2^{k+m}$个位图中选择出2^m个，让它们之间的距离尽可能得远。码字之间的距离必须达到最大可能的距离 d。

图4-20中的几何距离意味着两个点之间边的数量。这与两个点不同位的数量是相同的。例如码字110和101之间的距离为2。人们称这个距离为非欧几里德距离，将作为向量的码字所占的空间称为非欧几里德空间。欧几里德距离，即110和101之间的传统的几何距离是1.41。

距离 d 的含义是两个有效的码字之间的差距。汉明距离（Hamming-distance）h 是指以一个码为源的所有可能距离中的最小值，它是衡量一个码的有效性的尺度。

一个码字的权重是指在这个码字中出现的1的个数。

用数学方法可将一个码字描述为向量 $\boldsymbol{W}_i$，i 是索引下标。两个码字之间的距离 d 及一个码的汉明距离如下：

$$d(\boldsymbol{W}_i;\boldsymbol{W}_j)=w(\boldsymbol{W}_i\oplus\boldsymbol{W}_j) \tag{4-14}$$

$$h=Min\{d(\boldsymbol{W}_i;\boldsymbol{W}_j)\} \quad i\neq j \tag{4-15}$$

图4-20所示码的汉明距离为 $h=2$。我们现在应该做的是增大汉明距离。为了保留多种不同的描述方法，我们仍然设 $n=3$，$m=1$ 及 $k=2$。因此只有两种可能的码字，我们取000和111。图4-21在非欧几里德三维空间中给出了这些码字。

这两个码字构成了一个空间的对角。一个比特差错可能将111变成例如101。接收到的码字与111的距离为1，与000的距离为2。因为假设的是只有一个误码，因此发送的比特只能是111。

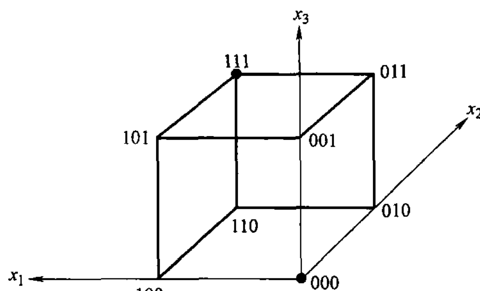


图4-21 汉明距离 $h=3$ 的码

$h=3$ 时单个比特的误码是可以纠正的。通过实验可以很容易地确定，$h=4$ 时也可对单个比特误码进行纠正，而 $h=5$ 时就可以对两个比特的误码进行纠正。一般来说有下面的结论：

在一个码字中可以对 e 个误码进行纠正，如果满足：

$$h\geqslant 2e+1 \tag{4-16}$$

在一个码字中可以识别出 e 个误码，如果满足：

$$h\geqslant e+1 \tag{4-17}$$

可以这样理解几何距离，空间中的每个有效点构成了一个可以纠错的范围。每个单个的范围不允许被打扰。解码器需要确认，收到的码字是属于哪个范围的，据此来确定它是哪个有效码字。

例如如果想要对5Bit的误码进行纠正，每个数据块需要11位的校验码。但使用11位的校验码可以识别10Bit的误码。即识别要比纠错的复杂性小很多，正是由于这个原因ARQ系统比FEC系统要多很多。FEC系统只应用在没有反馈信道（或存储中）的地方，此时多个用户同时从一个发送者接收（如无线广播）。

如果有用位的数量 m，纠错能力 e 以及必需的汉明距离为 h 是已知的，那么就能确

定最小的校验位数 k：在可纠错范围有$\binom{n}{1}$个单比特误码，$\binom{n}{2}$个双比特误码及$\binom{n}{e}$个 e 比特误码等。共有 $\sum_{i=1}^{e}\binom{n}{i}$个可以纠错的码字。纠错范围涉及正确的码字。因为$\binom{n}{0}=1$，所以在纠错范围内共有 $\sum_{i=0}^{e}\binom{n}{i}$个码字。总共有 2^m 个纠错范围，即最少包括 $2^m\cdot\sum_{i=0}^{e}\binom{n}{i}$个不同的比特组合。但最大的比特组合数为 2^n，因此：

$$2^m\times\sum_{i=0}^{e}\binom{n}{i}\leqslant 2^n=2^{m+k}=2^m\cdot 2^k$$

$$\sum_{i=0}^{e}\binom{n}{i}\leqslant 2^k$$

$$k\geqslant\log_2\sum_{i=0}^{e}\binom{n}{i}=\log_2\sum_{i=0}^{e}\binom{m+k}{i}\tag{4-18}$$

上式可使用迭代的方法进行求解。

R_C 的值称为编码率：

$$R_C=\frac{m}{n}=\frac{m}{m+k}$$

R_C 也是数据量放大的一种度量方法。由于信道编码数据传输速率被增大到 R/R_C（或者说在信道数据传输速率相同的情况下，有用数据的吞吐量会下降）。R_C 的值小于 1，在实际中 R_C 在 0.5 和 1 之间。

当 h 较大时，由信道编码引起的数据量会急剧增大，除非 m 和 n 都很大。例如在航天飞行中码字的长度 $n=255$。在这样的空间中不可能通过实验得到 2^m 个汉明距离为 h 的码字。因此编码和解码都需要一个计算规则。当使用特殊的码组时这个规则当然会很简单。在实际中主要使用线性码组，特殊情况下会使用它的一个子组，称为循环码，如图 4-22 所示。编码要么使用矩阵计算方法要么使用多项式代数方法。

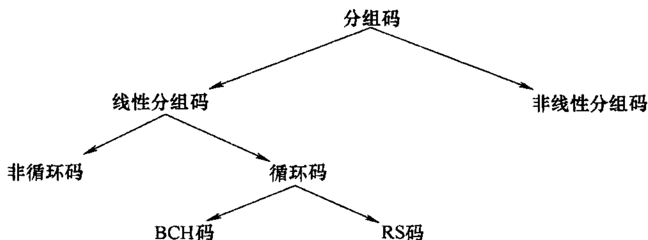


图 4-22 分组码的分类

其中的 RS 码（Reed-Solomon-Code，发明者的姓名）应用在 CD 中，而 BCH 码（Bose Chaudhuri Hocquenghem，发明者的姓名）在 DAB（Digital Audio Broadcasting，数字 UKW 无线广播）中得到应用。

在线性码中异或和（任意两个有效码字按位进行模 2 相加）还是一个有效的码。因此有 n 个 0 的位图也是一个有效码字，它是一个码字与自身相加的产物。线性码的汉明距离 h 相当于所有码字的最小权重，除非这个码字都为 0。

系统码是分组码，它的码字按照图 4-19 所示的结构组成，即首先是 m 个有用比特，然后是 k 个校验比特。比特的交织是可以想像的，但在实际中很少用在系统码中。所有的线性码都能用系统的方式进行描述。

使用 2 个数字来描述分组码：(n, m) 码

这里　n = 包括校验比特的码字的长度

　　　m = 每个码字中有用比特的个数

给定 n 和 m 的汉明距离的上限是不能够准确计算出来的。在文献中给出的值各不相同，如 Hamming 极限、Plotkin 极限、Singleton 极限等。

Plotkin 极限：
$$h \leqslant n \cdot \frac{2^{m-1}}{2^m - 1} \tag{4-19}$$

式 (4-19) 是个近似值，对二进制码并且当 m 的值较大时是适用的。14 位的码字，其中的 8 位为校验比特预留，那么可以构成汉明距离为 7 的码。而 8 位码，4 位为校验位，汉明距离可以是 4。根据式 (4-16)，可以对 1 位误码进行纠正。与在上节中介绍的带有奇偶位的分组可靠性中给出的例子具有相同的结果。

现在我们来看一下线性 (6, 3) 码是如何构成的。这个码共有 $2^m = 2^3 = 8$ 个长度为 $n = 6$ 的字，其中 $n - m = k = 3$ 位用作校验。汉明距离 h 为 3，根据式 (4-19) 这是可能的。

用向量来描述有用数据字 ($m = 3$ 个元素)：$\underline{d} = (d_1 d_2 d_3)$

用同样的方法来描述码字 ($m = 6$ 个元素)：$\underline{c} = (c_1 c_2 c_3 c_4 c_5 c_6)$

现在构成所谓的生成矩阵 (Generator matrix)。它必须是 $m = 3$ 的线性不相关行。需要对 $\boldsymbol{G}$ 进行变换，使它的前 m 列和 m 行构成一个单位矩阵。这种变换始终是可行的，它将线性分组码变成系统分组码。这样 $\boldsymbol{G}$ 就可以分成一个单位矩阵 [$\boldsymbol{E}^m$] 和一个行数为 m，列数为 $n - m = k$ 的矩阵 [$\boldsymbol{G}'$]。对我们这个例子来说，

$$[\boldsymbol{G}] = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$
$$\underbrace{\qquad\qquad}_{[\boldsymbol{E}^m]} \quad \underbrace{\qquad\qquad}_{[\boldsymbol{G}']}$$

关于码字的所有信息都在子矩阵 [$\boldsymbol{G}'$] 中。

现在根据规则构成码字 $\underline{c}$：
$$\underline{c} = \underline{d} \cdot [\boldsymbol{G}] \tag{4-20}$$

所有的计算都是模 2 的算术运算。

我们这个例子的结果是：

数据字：	码字：
000	000000
001	001101
010	010011
011	011110
100	100110
101	101011
110	110101
111	111000

在接收端准备一个所谓的控制矩阵［$\boldsymbol{H}$］：

根据 ［$\boldsymbol{G}$］ $=[\boldsymbol{E}^m \boldsymbol{G}']$ 可得 ［$\boldsymbol{H}$］$=[\boldsymbol{G}'^{\mathrm{T}} \quad \boldsymbol{E}^{n-m}]$ (4-21)

在我们这个例子中［$\boldsymbol{H}$］为

$$[\boldsymbol{H}]=\begin{bmatrix}1&1&1&0&0&0\\1&1&0&0&1&0\\0&1&1&0&0&1\end{bmatrix}$$

$$\underbrace{\quad\quad}_{[\boldsymbol{G}'^{\mathrm{T}}]} \quad \underbrace{\quad\quad}_{[\boldsymbol{E}^{n-m}]}$$

实际的校验操作由下面的乘法构成：

$$\underline{s}=\underline{c}\cdot[\boldsymbol{H}]^{\mathrm{T}} \tag{4-22}$$

向量$\underline{s}$的长度为 $m=3$，它被称为症状向量。如果传输中没有误码，那么$\underline{s}=0$。对我们的例子来说：

$$\underline{s}=\underline{c}\cdot\begin{bmatrix}1&1&0\\0&1&1\\1&0&1\\1&0&0\\0&1&0\\0&0&1\end{bmatrix}=(0\ 0\ 0) \quad \text{对所有的有用码字而言}$$

如果校验运算结果表明，误码的比特数与可纠正的比特数相等，那么可以根据症状向量 $\underline{s}$ 计算出误码向量，这个向量能给出错误的，只需要进行反转的比特。使用 Berlekamp-Massey 算法能对症状向量进行评估。该算法是由 Berlekamp 开发的，但 Massey 第一次给出了对该算法的可行性的解释。

线性码和系统码的一部分，即循环码的计算比较讲究。它基于多项式代数而不是矩阵计算。这种方法称为 CRC（Cyclic Redundancy Check）码。由于两个原因，该算法得到了广泛的应用：一是它能识别突发误码，二是多项式计算比矩阵计算更简单、更快。

此时数据字不被看成是向量，而是被看成多项式。每个单个比特正好是多项式的系数。例如，数据字

$$I=(1\ 1\ 0\ 1\ 0\ 1\ 1)$$

相当于下面的多项式：

$$I=1\cdot b^6+1\cdot b^5+0\cdot b^4+1\cdot b^3+0\cdot b^2+1\cdot b^1+1\cdot b^0=b^6+b^5+b^3+b^1+b^0$$

生成矩阵由生成多项式 G 来代替。它的阶数为 k，使用它可以构成 k 个校验位 P。它是由除法的余数构成的

$$\frac{b^k I}{G}=Q+\frac{P}{G} \tag{4-23}$$

I 是包含 m 个有用比特的多项式。它与 b^k 的乘积的含义是向左移 k 位并在末尾补充 k 个 0。数字信号处理中的 Z 变换的含义与它正好相同：只不过那里通过与 z^{-1} 相乘，多项式是向右移的。因此在发送端中构成了码字：

$$C=b^k I+P \tag{4-24}$$

它的含义是 m 个有用比特向左移 k 位，然后在末尾加上 k 个校验比特。

接收端将接收到的码字除以 G。因为除法 I/G 的余数在发送端与 $b^k \cdot I$ 相加，因此如果没有传输误码的话，在接收端该除法就不会有余数。

发送端和接收端的运算是相同的，可以使用移位寄存器简单、快速地实现。

循环码这个名字来源于对一个码字进行循环位移又能产生另一个码字的事实。因为是线性的，所以所有的码字都能从一个位图通过位移和叠加来产生。这个位图就是上面提到过的生成多项式，当然它不是完全随意选择的。可以使用所谓的原始多项式，它的系数在表中列了出来。

CRC 码有一个最好的特性，就是 k 个校验位能发现 k 个相邻的误码（突发误码）。因此 CRC 码已经成为一个通常的采用下面生成多项式的事实上的标准：

CRC-16：$G(x) = x^{16} + x^{15} + x^2 + 1$

CRC-CCITT：$G(x) = x^{16} + x^{12} + x^5 + 1$

例如：

有用字：　　　　　　　　I = 0 1 1 0 0 0 0 1 1 0 1

生成多项式的系数：　　　G = 1 1 0 1 0 1

通过连续的除法并利用余数可以计算出校验位：

```
1 0 1 0 0 0 1 1 0 1 0 0 0 0 0  （给 I 添加 k=5 个 0）
1 1 0 1 0 1
  1 1 1 0 1 1
  1 1 0 1 0 1
      1 1 1 0 1 0
      1 1 0 1 0 1
          1 1 1 1 1 0
          1 1 0 1 0 1
              1 0 1 1 0 0
              1 1 0 1 0 1
                  1 1 0 0 1 0
                  1 1 0 1 0 1
                        1 1 1 0 = 余数→P = 0 1 1 1 0 （k=5 位）
```

因此，得到的码字为：C = 1 0 1 0 0 0 1 1 0 1 0 1 1 1 0

在上面的两个例子中使用的都是比较熟悉的多项式计算和矩阵计算，唯一不同的是在所有的运算中都采用的是模 2 的算术运算。下面我们仔细分析一下这种非常规的计算方法。

我们已经习惯用强大的代数运算来求解复杂的任务，绝大多数情况下是对实数进行运算，有时也应用于复数。但将这么强大的代数运算用到编码技术上还有些诱人。

传统的代数被定义成数域。数域是具有一定属性的数据集。主要属性包括：

1）域对加法来说是闭合的：一个数域中的两个数相加，结果仍然属于这个数域。

2）域对乘法来说是闭合的。

3）对加法来说存在一个中性元素，一个数与中性元素的和等于这个数本身。

4）对乘法来说也存在一个中性元素。

5）分配率成立，即 $a(b+c)=ab+ac$。

6）除法必须重新定义。

如果定义其他的属性，那么就是组或环。数域的例子包括实数集和复数集等。0 和 1 是中性元素。域包括无穷多的元素，否则的话它就不是闭合的。因此数域也被称为无穷域。

同一个代数可应用于所有的数域，因此计算规则对实数和复数是相同的。实数可以用连续数产生的图形来描述。

用实数可以产生多项式。现在我们看一下不可约分的多项式，也就是说多项式的根不是数域的元素。例如方程 $x^2+1=0$ 在实数域内无解。一种解决办法是对数域进行扩展，这样实数就扩展到复数。图形描述需要有另一个维度，因此人们采用了数平面的概念。

字长为 n 的二进制数（这样可以概括所有的数字数据）只能描述 2^n 个不同的数。使用这样的数据字不能构造一个无穷域。因此采用一些技巧：不使用数线，而使用数环的概念。例如，一个有限域包括 1 位的十进制数 0～6。如果将两个数相加，如 3+5，那么在数线上得到的结果是 8。但这数环上得到的结果是 1。这无非就是模 7 的运算，这里 7 是有限域内的一个数元素。

但是有限数域如果不打破上面的规则的话，是不能用任意的数元素来构成的。Franzose Galois 已经证明，有限域中的数元素必须是质数或质数的幂（顺便提一下 Galois 21 岁时死于一场决斗）。因此人们称他为 Galois 域 GF（p）及 GF（p^n）。这里 p 是一个质数，n 是自然数。在英文中域用 field 表示，Galois 域为 Galois-Field，它的简称即为 GF。在德文文献中也经常用 Galois-Feld 来表示 Galois 域。

可以证明 GF（p^n）是 GF（p）的一个扩展。对具体的应用还需规定 p 和 n 的值。P 是指数，最小值是 $p=2$。Galois 域 GF（2）包含元素 0 和 1，它们即是中性元素又是可以表示比特的逻辑值。Galois 域 GF（2^n）包含 2^n 个元素，n 正好是数据字的长度（位数）。

使用 Galois 域的独特技巧在于将很熟悉的代数方法应用于面向数字信号的数域中。

GF（2）中的加法有下面几种情况：$0+0=0$

$$0+1=1$$
$$1+0=1$$
$$1+1=0$$

这同模 2 的加法没有任何区别，也相当于进行异或（EXOR）运算。很明显中性元素 0 的作用是可以看得出来的。根据对数环的理解可以得出这样的结论：在 GF（2）中加法和减法是相同的。

GF（2）中的乘法有下面几种情况：$0\cdot 0=0$

$$0\cdot 1=0$$
$$1\cdot 0=0$$
$$1\cdot 1=1$$

这相当于与（AND）运算。

BCH 码和 RS 码（参见图 4-22）都使用 Galois 域的代数寄存器。它们甚至在 GF(2^n) 中使用离散傅里叶变换（DFT）！这点可以解释得很清楚：

我们先回到实数来。遵照采样定理对时间受限的信号进行采样，会得到 m 个实数值并将它们存储起来。可以对这 m 个数进行 DFT，可以产生 m 个频域内的值。

现在对同样的信号再次进行采样，但是稍快些。因此会产生 $n>m$ 个采样值，然后也对它们进行 DFT。这次会产生 n 个频域值，但最上面的 $k=n-m$ 个频域值都为 0，因为时域信号采样过快。

现在我们的想法是这样的：将 m 位有用数据字取出。这些数据字被认为是频域向量，然后加上 k 个 0，将它们的长度扩展成 n。经过反 DFT 运算会产生时域向量，这些 0 就会分散开来而不能显示观察到。时间序列的长度为 n，因此这相当于过采样的信号，时域的值是冗余的。现在这个时域序列就是码字，需要向接收端传输。这会产生一个 DFT，并且需要检查最高的 k 个频域值是否为 0，这样就是症状向量。如果在传输的过程中只有一个采样值发生错误，在症状向量中就会有指示。DFT 是一个 N 到 N 的变换，即采样值会影响每个频域的值，反过来也是这样。图 4-23 给出了这个过程的原理。

BCH 码和 RS 码的区别在于，BCH 码使用的是二进制码（二元码），而 RS 码使用的是多元码，绝大多数情况下码元数量是 2 的指数倍。根据观察问题的角度，RS 码可以是 BCH 码的一个特例，BCH 码也可以是 RS 码的一个特例。

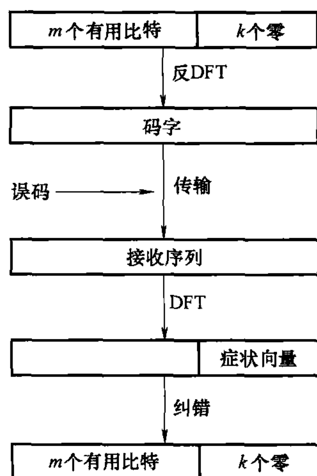


图 4-23　BCH 码和 RS 码的转换编码

在数据传输过程中，接收器一直在不断地作为级联的进判决-线路解码器来起作用，如图 2.11 所示。如果引入信道编码的话，信道编码会与线路编码放在一起，在带通传输中，会将解调器放在前面。定时采样和解码是有区别的两种方法：

1）硬判决解码：信道解码器根据判决器后的逻辑值，使用最小非欧几里德距离对码字进行纠正。

2）软判决解码：信道解码器根据采样时刻的实信号值（在判决器之前），使用最小欧几里德距离对码字进行纠正。

例如：码的内容为 1 1 1 和 0 0 0。在双极性传输中，这些字分别使用 1、1、1 及 −1、−1、−1 的电平来表示。接收到的并且经过采样的在判决器之前的电平值为 0.5、0.5、−3。

硬判决解码器首先对一个个比特进行量化，送给信道解码器的值为 1 1 0。信道解码器将它纠正为 1 1 1。

点（0.5 0.5 −3）与点（1 1 1）之间的欧几里德距离的二次方是 $(1-0.5)^2+(1-0.5)^2+[1-(-3)]^2=16.5$，与点（−1 −1 −1）之间的距离会小些，为 $0.5^2+0.5^2+2^2=8.5$。因此软判决解码器给出的结果是 0 0 0，即与硬判决解码器给出的值刚好相反。这是

因为第三个比特偏离的距离很大。当偏离的比较小时两种解码器会给出相同的结果。

当然软判决解码比较好，但它较复杂因此用得较少。多数情况下在接收端使用的不是像图 2-13 所示的那样的简单的比较器，而是像图 2-14 所示的 A/D 转换器。因此硬件对软判决解码器比较有利。

可以通过对码进行交织的办法提高可靠性，如图 4-24 所示。例如在 CD 中就对两组 RS 码进行了交织。

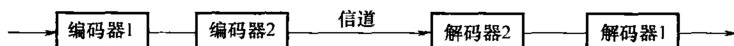


图 4-24 经过交织的码（链接码）

当在一个块中出现太多的误码时每种纠错方法都会不起作用。当出现突发误码时，就会出现这种情况，尽管此时 BER 仍然会很小。也就是说，绝大多数的码字包括太多的冗余，而单个码字的冗余却太少。总的来说这种码并不是很有效。

针对这种情况一种改进方法是采用交织的办法对已经编码的比特进行扰乱。这要求多个块同时进行，相应地会引起一些传输时延。交织（码展开）是一个简单的操作：将已编码的块按行读入一个矩阵，然后按列读出。接收器执行相反的操作。长度为 e 的突发误码经过这样的变换会转成 e 个单独的误码。汉明距离 $h=3$ 就足以对这些单个误码进行纠正。这样做是以牺牲电路的复杂性和时延来节省带宽。CD 中光盘表面的污染经常会导致突发误码，因此在 CD 中也使用交织并不奇怪。如 4-25 所示为使用交织器的结构图。

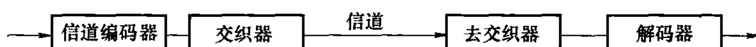


图 4-25 利用交织来解决突发误码

4.3.3 卷积码

卷积码（convolutional codes）组成两组信道码。它不是按块，而是连续进行编码。卷积码的理论还比较年轻，与分组码相比它还不是很明了。尽管如此，卷积码很容易实现，使用软判决解码的 Viterbi 算法是一个很强大的解码方法。Viterbi 算法的一个变种顺序解码（Stack-Decoder 和 Fano-Decoder）也同时被使用。卷积码是线性的，但不是系统的。

卷积码从移位寄存器中按行读取有用数据。读出数据后进行适当的数学运算，当然是模 2 的算术运算，结果是产生校验比特。图 4-26 给出的是一个简单的例子。

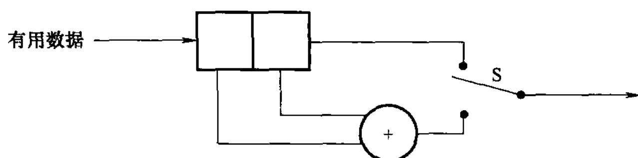


图 4-26 Wyner-Ash 码的卷积编码器

这里的开关 S 按有用数据时钟的两倍的频率上下来回变换，在输出端包含原始数据，但每两个比特之间包含它们之间的异或值。编码之后的数据量也变成了原来的两倍。从下面的有用数据中

B1　　B2　　B3　　B4　　B5　　等

会产生下面的编码数据

B1　　B1 ⊕ B2　　B2　　B2 ⊕ B3　　B3　等

接收端的解码器重复同样的运算并将运算的结果与收到的校验比特进行比较。如果出现单个的误差，则校验比特在传输中出现误码。如果两个相邻比特的运算产生不同的结果，那么这中间的有用比特出现误码。

当在两个误码之间（包括校验比特）至少有 3 个正确传输的比特时，Wyner-Ash 码会发现所有的误码。因此 Wyner-Ash 码可将误码从 25% 降低到 0%，付出的代价是数据量翻倍。

在 4.3.2 节中作为例子介绍的（6，3）分组码汉明距离 $h = 3$ 时每个码组使用 6Bit 才能对 1 个比特进行纠正。尽管数据量也翻倍，最大误码率为 16%。

当然在图 4-26 中编码器使用的移位寄存器也很长。此时一个比特的信息被分散在多个比特中。复杂的码并不像分组码那样直接进行计算，而是采用仿真的办法找到一个最优码。图 4-27 所示为一种较复杂的卷积码的结构。它每次将 k 个有用比特推入长度为 $L \cdot k$ 的移位寄存器中，然后对寄存器中的内容进行 n 线性组合。此种方法的编码率为 $R_C = k/n$。

通常情况下卷积码的编码率较小，即有用数据的吞吐量很差。相比之下这种方法的纠错能力过强，因此人们通常去掉几个比特，事先造成几个比特的误码，解码器可以毫不费力地将它们纠正过来。这种所谓的收缩码能够提高有用数据的吞吐量。

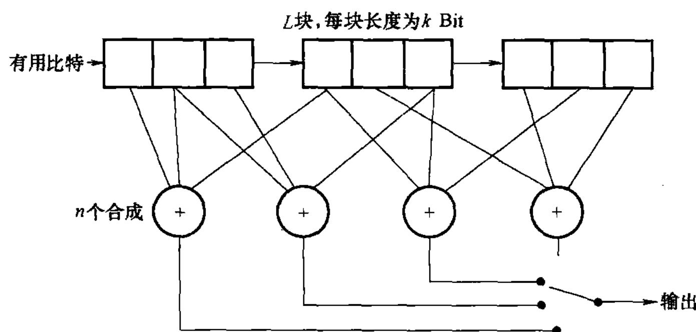


图 4-27　卷积编码器的通用结构

卷积码可以理解为自动机，移位寄存器的瞬时值表示的是当前的状态。它也能描述自动机的历史值。在卷积码中使用的不是在自动机理论中常用的、能够清晰描述状态变化的状态图，而是在时间轴上描述不同的状态。图 4-28 描述的是使用两个移位寄存器单元具有 4 个状态的卷积编码器。

图 4-28 中 4 种状态之间的 3 个区域是相同的，它表示的是同一个自动机。每个状

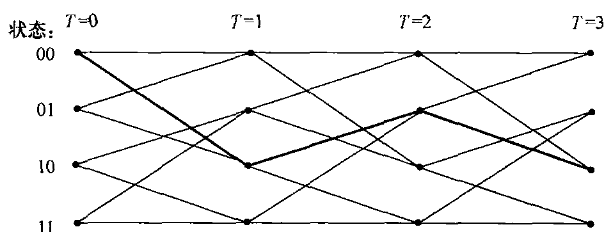


图 4-28 卷积编码器格子图

态有两个出口，它们与下一个比特是逻辑 1 或逻辑 0 无关。出口的可能目标与卷积编码器的异或组合有关。图 4-28 很像一个格子（英文为 trellis），这也是格子图这个名称的由来。图中粗线表示的是一个确定的数据流在格子图中的路径。如果遇到传输差错，路径就会出现偏转。Viterbi 在 1967 年发表一种有效的算法，它能给出格子图中各种可能路径的概率。

4.3.4 编码方法的选择

没有一种编码方法是理想的，否则在前面的几个小节中就介绍了。通常情况下是根据应用来选择一种最优的编码方法。因此像选择调制方法一样，要进行折中。

选择编码方法时要考虑的因素包括：解码器后可接受的误码率、信道干扰的特性、数据传输速率、信息的类型、未被发现的差错或被错误纠正的差错所引起的后果、带宽增加和设备复杂性所带来的开销和可接受的附件时延。

所引起的开销需要与采用其他类型的方法所带来的复杂性，例如提高发射功率，进行对比，人们一直在不断寻求一种整体最优的方法。值得注意的是，需求说明不能是简单地要求一个没有差错的系统，但也不能模糊地表达要求一个“尽可能好的系统”。

通常情况下 BER 为 $10^{-5} \sim 10^{-7}$ 的 AWGN 信道（加性高斯白噪声）采用使用 Viterbi 解码器的卷积码是合适的。但在这种信道中会出现单个误码。

如果信道的误码不具有高斯特性，那么会出现突发误码。此时采用 RS 码或 BCH 码的效果会较好，特别是要求解码后 BER 小于 10^{-10} 时。同样在 FEC 系统中适合使用循环码。

但 BER 在很多情况下是值得怀疑的质量评价标准。特别小的 BER 通常会有波动效应，例如在计算机程序中无差错是很重要的，但在一个程序中出现一个差错还是出现一百个差错是无所谓的，因为不管是哪种情况程序都是不可用的。

根据信息源的不同情况，对所有的比特都进行保护也是没有意义的。例如在 PCM 信号中只对最前面的比特提供冗余是很有效的，因为经过翻转的 LSB 几乎是不能感知到的。因此这里 BER 不是最主要的。MR-QAM 也需要进行同样的考虑，如图 3-47 所示。

同样对 PCM 信号也不需要一直进行纠错。如果一个采样值传输发生差错，可以很方便地采用对相邻采样值进行差值的办法来解决。人们将这种技术称为差错覆盖或差

错掩藏。在 CD 中也会使用这种技术。

差错掩藏在所有的主观评价系统（即信宿是人而不是机器）的传输中都是适合的。但信源信号压缩得越厉害，允许的误码就越小。

理论研究通常都是以均匀 AWGN 信道为基础。均匀是指传输的 0 的个数与 1 的个数平均起来是相等的，并且 0 被误传成 1 及 1 被误传成 0 的概率都是相同的。TTL 信道不是均匀信道，因为 1 比 0 多。非均匀信道通过使用一个附加电路（编码转换）会变成均匀信道。

无记忆信道有一个固定的误码概率，有记忆信道会产生突发误码。

在信道编码之后还会进行调制，因此必须考虑它对差错特性的影响。例如调制前的差分编码（目的是简化解调中的同步）容易引起双比特误码。

如果分组编码中的组长度选择得较小，控制比特与有用比特相比占的比例会很大，有用比特的吞吐量会下降。如果组长度选择得很大，分组编码的误码率会增大（参见 4.3.5 节）因此在 ARQ 系统中重发的数量会增加，有用数据的吞吐量也会下降。可以计算出最优的分组大小，以保证在误码率相同的情况下有用数据的吞吐量最大[Ger94]。但是在实际中必须保证信道的特性是已知的或是可以计算的。系统使用一种简单的、很实际并且很有效的办法：当值得怀疑的重传的数量增大时，就减少组的长度，否则就增加组的长度。这涉及到一个可适配的协议。

由此可以得出一个结论，即应该尽量使用 ARQ 方法，只不过需要实现一个可适配的协议。同样针对有用数据吞吐量，ARQ 系统也应该考虑到 FEC 系统的原理：

1）FEC 系统中信道编码会特意增加很多冗余。如果没有遇到干扰，所附加的冗余毫无用途，有用数据传输速率较低。

2）相反 ARQ 系统使用较小的冗余进行分组，只重传被破坏的组。这种重传同样是冗余，但只是在必要的情况下才进行。如果受到的干扰很强烈，有用数据的吞吐量同样会很小，甚至下降到零。经过多次无效尝试后连接自动中断，将信道空闲。

最后还值得一提的是，编码并不仅仅是为了误差保护，也可用于同步。失去同步的症状是系统持续不断地报告传输差错。通常情况下传输带宽用于：

1）传输真正有用的信息。

2）通过调制和信道编码进行误差保护。

3）传输附加信息，如像导频和同步信息那些只供解调和解码使用，而并不传送给信宿的信息。

人们将这些任务结合起来，因此可以节省一些带宽。

4.3.5 编码增益

信道编码会给整个传输系统带来哪些效益？乍看起来答案很简单：通过传输冗余来减少误码率。但这些冗余会需要传输带宽或传输时间或存储空间。对编码方法进行评价的前提假设是传输时间相同。信道编码引起的数据传输速率的翻倍对带宽的需求也增加了。这在 AWGN 信道中会引起噪声功率的增加，即根据式（1-20）信道中的信噪比会减小，误码率会增加。信道编码必须对这种提高进行补偿，否则信道编码所带

来的复杂性是不值得的。

香农已经证明了只要信息速率 J 小于式（1-17）中给出的信道容量 C，在 AWGN 信道中无差错传输就是可能的。但香农并没有给出什么样的信道编码会实现这个边界值，他只是证明了这样的编码是存在的。在有些情况下这样的编码是很复杂的，并不能在实际中应用。

这涉及的并不是如何达到理论上的极限值，而是与没有采用编码的传输情况相比情况会有所改善。这种改善可以使用两种等价的方式进行描述：

1）在相同的发送功率情况下，通过编码会使误码率下降。

2）在相同的误码率下，可以减少发射功率。

如果使用第二种定义方法，尽量减少发射功率，这样编码增益可以与调制增益进行比较。例如在图 2-24 中当 BER 为 10^{-10}时采用双极性传输编码增益（通过 BCH 码实现）约为 4dB。此时由于编码而引起的带宽增加从而导致 SR_K 的降低已被考虑在内。

正如已经说过的在实际中没有不出现误码的情况。为此我们以分组码信道编码为例计算剩余误码率。信道采用无记忆的 AWGN 信道。

某一个比特在传输中会出现错误的概率为 p_{Bit}，被正确传输的概率为 $1-p_{\mathrm{Bit}}$。连续 m 个被正确传输的比特的概率为 $(1-p_{\mathrm{Bit}})^m$，因为信道是无记忆的，所以各个比特之间是无关的。

这也可以用另一种方式表达：长度为 m 个比特的比特组出现了 i 个比特的误码，即被正确传输的比特数为 $(m-i)$，这种情况的概率为

$$p_{\mathrm{Bit}}^i(1-p_{\mathrm{Bit}})^{m-i}$$

因为从 m 个比特中选出 i 个错误比特总共有 XX 种可能性，长度为 m 个比特的比特组中出现 i 个错误比特的概率为

$$\binom{m}{i}p_{\mathrm{Bit}}^i(1-p_{\mathrm{Bit}})^{m-i} \tag{4-25}$$

令 $i=0$，由于 XX $=1$ 即可得到整个比特组都被正确传输的概率为 $(1-p_{\mathrm{Bit}})^m$，与上面的结果是一致的。

一个长度为 m 的比特组出现误码的概率，即比特组至少出现 1 个比特误码的概率称为块误码概率 p_{Block}，它的值为

$$p_{\mathrm{Block}}=1-(1-p_{\mathrm{Bit}})^m \tag{4-26}$$

这个值也可以采用另一种方法来计算。这种计算方法比较复杂，但后面我们还需要这个另外一种形式的结果。

块误码概率是出现 1 个或者 2 个或者 3 个…或者 m 个比特误码的概率，根据式（4-25）它的值为

$$p_{\mathrm{Block}}=\sum_{i=1}^{m}\binom{m}{i}p_{\mathrm{Bit}}^i(1-p_{\mathrm{Bit}})^{m-i} \tag{4-27}$$

式（4-26）与式（4-47）表示的内容是相同的，因此它们一定是相关的。这可以用二项式定理来解释

$$(a+b)^m = \sum_{i=1}^{m} \binom{m}{i} a^{m-i} b^i$$

现在设 $a = 1 - p_{\text{Bit}}$，$b = p_{\text{Bit}}$：

$$(1 + p_{\text{Bit}} + p_{\text{Bit}})^m = 1 = \sum_{i=1}^{m} \binom{m}{i} p_{\text{Bit}}^i (1 - p_{\text{Bit}})^{m-i} \tag{4-28}$$

与式（4-27）相比，它的唯一区别在于和的表达式是从 0 开始的。因此应该将 $i=0$ 的值从中减掉。

$$p_{\text{Block}} = 1 - \binom{m}{0} p_{\text{Bit}}^0 (1 - p_{\text{Bit}})^m = 1 - (1 - p_{\text{Bit}})^m$$

因此得到了与式（4-26）相同的结果。

$$p_{\text{Block}} = \sum_{i=1}^{m} \binom{m}{i} p_{\text{Bit}}^i (1 - p_{\text{Bit}})^{m-i} = 1 - (1 - p_{\text{Bit}})^m \tag{4-29}$$

根据式（4-24），p_{Bit}可以用信道的信噪比 SR_K 来表示。图 4-29 给出的是当比特组的长度为 1，16 及 512 时 p_{Bit}值的情况。$m=1$ 时图 4-29 的结果与图 4-23 给出的误码率的曲线是一致的。

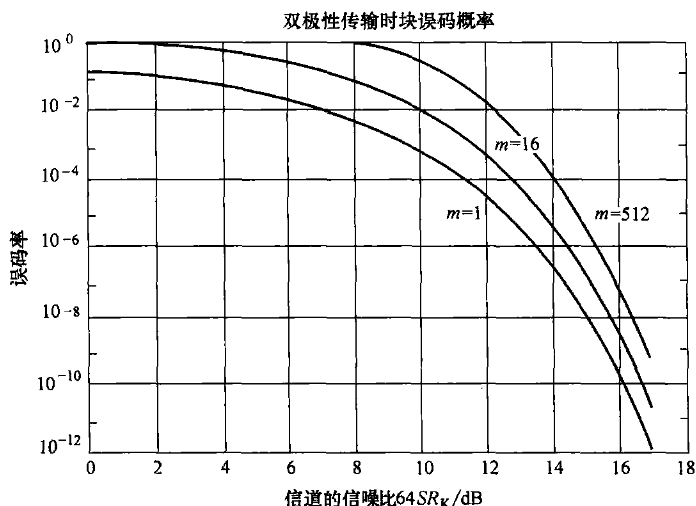


图 4-29　位组长度 $m=1$，16，512 时不可靠位组的块误码概率

从图 4-29 中可以看出，当 SR_K 为 8dB 时，长度为 512 的码组根本不可能进行无差错传输。

剩余误码概率（Rest）p_R 指的是引入信道编码之后的块误码概率。在长度为 $n = m + k$ 的码组中最多出现 e 个误码的概率可以根据式（4-29）得到，其中包括不出现误码的情况，即其中的和式是从 0 开始的，而不是从 1 开始：

$$\sum_{i=1}^{e} \binom{n}{i} p_{\text{Bit}}^i (1 - p_{\text{Bit}})^{n-i} \tag{4-30}$$

解码器最多能纠正 e 个误码，多于 e 个误码的位组不能被纠正。这个概率正是我们

已经求得的剩余误码概率 p_R，根据式（4-30），这个值可以很容易的得到：

$$p_R = 1 - \sum_{i=1}^{e} \binom{n}{i} p_{Bit}^{i} (1 - p_{Bit})^{n-i} \tag{4-31}$$

根据式（4-18）最大可能纠错值 e 可以根据 m 和 k 得到。差错识别的剩余误码概率总是要小于差错纠正的剩余误码概率。

例：使用 BER 为 10^{-4}的电话线传输数据。为此将数据按照不同的方式进行分块，首先将它们分为 4 个没有保护的块，然后采用（7，4）的分组码。我们比较第一种情况下的块误码概率与第二种情况下的剩余误码概率。

第一种情况：根据式（4-29）给出的块误码概率与 $e=0$ 时式（4-31）给出的剩余误码概率相同，值为

$$p_{R1} = 1 - (1 - p_{Bit})^4 = 1 - 0.9999^4 \approx 4 \times 10^{-4}$$

第二种情况：根据式（4-19），（7，4）分组码的汉明距离 h 最大值为 3。根据式（4-16）它最多能够纠错的比特数 1，即 $e=1$。将它代入到式（4-31）中有

$$p_{R2} = 1 - (1 - p_{Bit})^7 - 7p_{Bit}(1 - p_{Bit})^6 = 1 - 0.9999^7 - 7 \times 10^{-4} \times 0.9999^6 \approx 2 \times 10^{-7}$$

简单的分组码能够将传输的可靠性提高 10^3，但是传输时间几乎翻倍。

编码增益 G_C 根据块误码概率与剩余误码概率的关系来定义。根据式（2-24）p_{Bit} 可以用 SR_K来表示

$$G_C = \frac{P_{Block}}{P_R} = \frac{1 - (1 - P_{Bit})^{n-k}}{1 - \sum_{i=1}^{e} \binom{n}{i} p_{Bit}^{i} (1 - p_{Bit})^{n-i}} \tag{4-32}$$

用一个例子可以清楚地解释这个增益的含义。长度为 $n = 8 \sim 128$ 的分组码中应该能够对 3 个误码进行纠正。根据式（4-18）可以计算出 k，然后将它代入到式（4-32）。图（4-30）给出了这个结果。

根据图 4-30，当误码率为 10^{-3}时，采用长度为 $n=128$ 的分块方法可将误码率减小为 1/10000。为此需要 19 个校验比特，在每块中的 128 个比特中只能传输 109 个有用比特。

通过计算来识别误码并不能得出固定的结论，因为式（4-18）对上面这种情况并不适用。人们一直采用误码纠错（FEC 系统）的办法对它进行估计并且确定，误码识别（ARQ 系统）的结果一定比误码纠错的结果好。

从上面计算出的编码增益我们可以看出，与信道的误码率相比，解码后有用数据的误码率下降了很多。为了使编码增益（以牺牲带宽为代价换取来的）与调制增益（同样也是由带宽换来的）联系在一起，采用标准化的描述方法是有很多帮助的。这点我们在 2.8 节的末尾已经提到过了，可将 BER 描述成对每个比特投入的能量的函数。我们在图 2-24 中已经进行了观察，那里为了对比观察采用的就是 BCH 编码的曲线。

如果要想通过信道编码来降低 BER，并且不增加传输带宽和时延，那么必须使用高效的调制方法，但这样又会降低 BER。在这种情况下还是存在一个比较好的系统，即它接近香农极限值。原因是通过信道编码进行的干扰保护比通过调制进行的干扰保护要好。基于这种考虑，在性能较好的数字传输中采用 32QAM 或 64QAM，即采用具有较少干扰保护的调制方法。但为此最好使用较好的信道编码方法来保证可靠性。

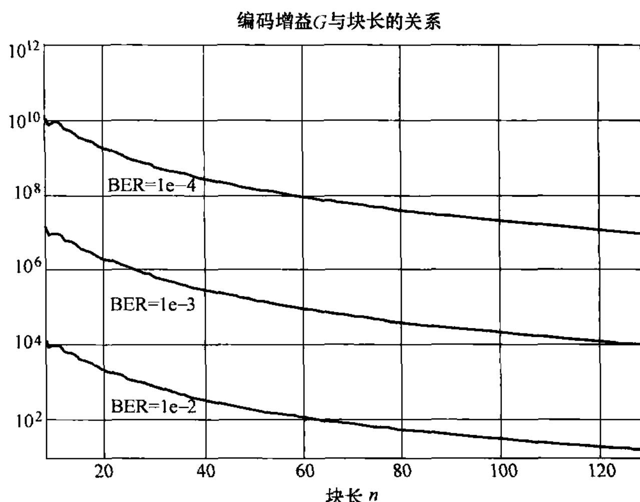


图 4-30　块长为 $n=8\sim128$ 及 3 种不同的 BER 情况下的编码增益

最重要的要将信道编码和调制作为一个整体来考虑。这会引出名为编码调制或网格编码调制这样的概念。

4.3.6　其他点到点的可靠连接

信息传输的基础是香农公式式（1-17）。它涉及的内容是如何将传输带宽和信噪比最大化。图 4-31 给出了一个概括的总结。

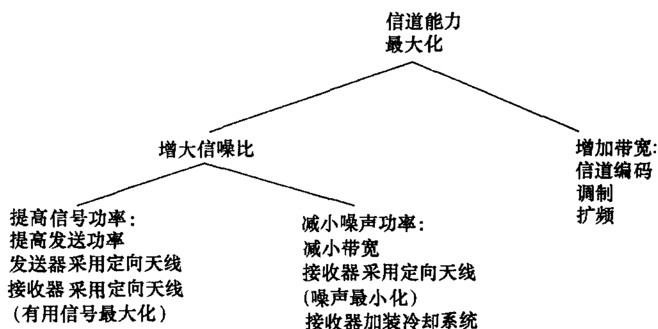


图 4-31　使连接更加可靠的方法

根据图 4-31，提高带宽 B 对信道的能力 C 有正、负两方面的影响。因此当 B 无限制增长时 C 会接近一个极限值，比较图 1-41。但随着功率的增加 C 会无限制的增加，比较图 1-40。

图 4-31 中提到的扩频方法会在 5.3 节中介绍。

我们在前面已经介绍过了信道编码和调制的主要方法，也提到了几种改进方法，如软判决解码、相关接收器、交织等。尽管如此还必须牢记：

> 绝对误码为零的传输是不存在的！采用足够复杂的措施可以将误码率任意地减小。

当然付出的代价是必须从技术上准确地确定，到底要求什么样的传输性能。

> 没有一种传输方法对所有的信道都是最优的。

因此必须对信道特性有所了解。如果信道特性发生变化，传输系统必须进行相应的调节，即选择不同的调制方法和信道编码方法。自适应系统可以自动完成这些功能，例如 ARQ 系统要求重传的次数、信道编码后的特有行为、（所熟知的）同步帧中的误码以及接收器输入端的功率等都可以是进行调节的依据。

自适应系统是应用最广的。因为接收器输入端的值起决定性的作用，而这个值受发送器发出的值的控制，因此自适应系统需要一个反馈信道。在这点上也建议使用 ARQ 系统。

通信对端可以通过 ARQ 证实信道得到通知，是否必须提高发送功率还是可以适当降低发送功率的值。受数字传输波动效应的影响，信噪比的值过大是没有意义的。

根据信道特性的不同，ARQ 系统的时延是可变的，因此它对具有实时要求的传输（例如话音）是不适合的。

除图 4-31 所述的方法之外，下面的方法也值得考虑：

1）分集运行：这不外乎就是一个处于运行状态的备份连接，有下面的几种情况：

频率分集：一个无线传输同时运行在两个频率上。两个连接同时被干扰的概率比其中的一条链路被干扰的概率要小很多。

空间分集：将连接进行空间上的分离。例如在 UHF 频段上的无线广播中经常会出现干扰的情况，即由于不同的反射而造成的严重的接收场中断。但距离天线几个波长远的地方，即几米的距离，场强却很大。因此采用空间上分离的两个天线所提供的连接比只使用一个天线可靠得多。

时间分集：一个消息可被重发一遍，目的是减少瞬时干扰。这种方法显然与信道编码和交织有许多共性。

极性分集：同时利用电磁波的水平极性和垂直极性（定向无线广播，参见 5.5.6 节）。

分集指一个有用的连接同时通过多条信道来进行。这与复用运行，即多个有用连接同时通过同一条信道来运行是不同的。但一个复用信号可以使用分集系统来传输。空分复用指与连接相关的导线对或同轴管被封装在同一个电缆外衣中（即处于同一条电缆中）而空间分集指连接新使用的导线对或同轴管处于不同的电缆中，这样挖掘机不能将两条电缆同时斩断。极性分集指通过水平极性和垂直极性信道传输相同的数据。而极性复用指的是传输不同的数据。

2）备用系统：该方法需要安装一套完整的备用传输系统。如果这套系统进入运行状态，它就是分集。如果这套系统已经接入，但没有运行，就是热备份。出现问题时就自动地没有时延地切换到这套备份系统中。人们也称之为（1+1）运行。备用系统与原系统的花费是一样的，因此人们可能对 5 个连接安装使用一套备用系统，即采用（5+1）的运行方式。备用系统也可能会出现设备的损坏，除此之外还需要进行不中断

有用连接的测量和维护工作。因此也需要强调对运行可靠性很重要的管理方面的工作。

备用连接和分集系统的利用率很低，它们以很大的代价来换取可靠性。如果不受一个连接的限制，那么就会有另外一种花费比较小的方案，即信息网或数据网。

网络同时为多个用户提供连接。每个用户的感觉是使用点到点的连接，但实际上他只是使用了一个逻辑上的连接，该连接在网络中有一条路径，如图 1-49 所示。这条路径从一个网络节点到另一个网络节点，即是一串的物理连接的组合。物理连接可以使用到目前为止的所有的方法。因为数字信号是可以再生的，因此一串物理连接包含多少个段是无所谓的，出现干扰时选择另一条路径即可。这相当于分集运行。但它有一个很好的特性，即逻辑连接 A 的备用信道同时可为逻辑连接 B 的使用信道。整个网络可同时为多个用户提供一个固定的吞吐量。这个吞吐量不用使用几个价格较高的物理连接，而使用多个价格低廉的物理连接和一个组合即可实现。同时也可以在网络中对负载进行均分。

3）存储 ARQ：正常的 ARQ 系统将错误的帧丢弃然后请求重传。存储 ARQ 可以对这种情况进行改进。假设要传输一个数据字，例如 00110011。如果出现一个误码，这个字就会被临时存储起来并要求重传，可能还会要求进行第三次重传。最终在临时存储器中可能会出现下面的值（例如）：

10110010

00010011

00110111

按列根据服从多数的判决原则可以得到字 00110011，结果也会使系统达到零误码，尽管对该数据字没有进行纠错传输。

4）设备的自检测（Built In Test Equipment，BITE）。将这种方法与远程监测和控制相结合，运维人员可以从管理中心对整个系统进行检测。但这要求传输能力的一部分是为系统管理预留（EOW，engineering order wire，业务信道）。

5）无间断供电：这是最贵的系统部件之一，因为根据需要可以安装一个蓄电池进行短时过渡或安装一个柴油驱动应急发电机进行较长时间的供电等。这些设备需要复杂的建设设施（储油灌室、通风设备等）。

第5章 传输系统和传输媒介

接下来的两章将要把到目前为止所介绍的概念综合起来形成一个用于信息传输的功能系统。这需要从一个新的角度对所提出的方案进行评价。到本章为止所使用的数学和物理基础或多或少都比较明确，即它们可用对或错这样的词来评价。即一个系统包含太多不同的方面，不太容易这样进行评价。在一个系统中总是存在对下面几个方面的折中：用户使用方面（如数据传输速率）、物理上和信息论上的限制（如香农极限）、技术上的限制（如计算能量、存储大小）、规定和标准、花费。

现在的问题不是一个系统是好或者是坏，而是它到底离最优工作状态有多近。而这个“最优”主要随着微电子技术的进步而改变，但已过时的系统仍然会运行很长一段时间。这是因为例如由于与现有的系统不兼容而引入一个全新的无线广播系统的花费是很大的。基于这种考虑，下面还要对已过时的模拟无线广播系统进行讨论。

5.1 无线广播技术

5.1.1 无线广播接收

1. AM 无线广播

人们根据发送天线发出的电磁波与地面之间的仰角来区分无线传播方式。短波情况下空间电磁波经过几百千米高的电离层的反射返回地面。长波情况下地面波沿着地球表面发生偏转。中波的特性是在白天类似于长波，而在夜晚类似于短波。正是由于这个原因这3个频段都能传输较远的距离。

电离层是分层的，每层都能将空间电磁波反射回来。电离层的状态以及每层的高度都是变化的，我们将在5.5节中对它的机制进行详细的描述。空间电磁波因此具有时变特性。表5-1对这些特性进行了总结。

表5-1 长波、中波及短波的传输特性

频段	无线广播的频率	使用的波	信道特性	距离
长波	148.5 ~ 283.5kHz	地面波	固定的	洲际的
中波	526.5 ~ 1606.5kHz	白天:地面波 夜晚:空间波	白天是固定的 夜晚是变化的	洲际的
短波	3.2 ~ 26.1MHz 又分成12个子频段	空间波	可变的,多径接收,多普勒效应	全球的

第一个无线广播发射器 1920 年开始运行。那时的技术只能使用带载波的 ZSB-AM。按照今天的观点来看这是一种开销很大的选择，我们在 3.1.3 节第 1 部分的结束部分就已经对比给出了评价。1995 年有超过 20 亿的 AM 接收器在运行，大约 20000 个发射站为它们提供服务，其中包括 2500 个短波发射站。

与 FM 相比，AM 是一种窄带调制方法。正是由于这个原因，AM 在其相应的频段内能使远距离传输成为可能，使用频分复用方法的、无数的发射器能够同时运行，正如 FM 能做到的那样。

在 AM 调制之前信息信号经过 200～4500Hz 的带通滤波器过滤。滤波器的下限能够改善对话音信号的理解，上限能够确保传输带宽并确定信道栅格。信息信号的带宽比电话系统稍微大一些，目的是使音乐的传输成为可能，当然对传输质量的要求没有像今天这么高，只是接近满意而已。

使用音节压扩器（参见 1.3.2 节）可以使信息信号的动态性减少 10dB，这样使用中等级别的调制深度即可，而覆盖范围也会增大。最大调制深度限制为 1，这样在接收器中使用价格比较便宜的、与包络解调器工作原理相似的非相干解调器即可。

信道栅格应该使载波的频率间隔保持在 10kHz（9kHz 的信号带宽加上由滤波器边缘造成的所谓的保护间隔）。尽管如此，绝大多数的发射器都设计成 9kHz 的栅格宽度并且允许由较窄的接收滤波器所造成的信号质量下降。在 MW 中，特别是 KW 中允许使用栅格外的频率。

尽管 ZSB-AM 有许多大家都熟知的缺点，但今天仍然在使用它。由于目前频率资源比较紧张，加上它覆盖的范围较大，因此这种窄带方法仍然在被考虑的范围之内。话音无线广播业务多年来一直使用 SSSC 调制。对无线广播来说这是不可能的，因为传输音乐要求频率必须确定，参见表 3-1。人们对引入载波抑制的 SSB 这个问题已经讨论很长时间了，但由于它与已经安装好的接收器不兼容，人们一直在犹豫是否需要换成它。后来关于这个问题的讨论完全停止了下来，因为有一种纯数字的解决方案很快就能实现。我们将在本节第 4 部分中解释这个概念。

2. FM 无线广播

FM 能以牺牲带宽为代价来提高抗干扰能力。在一个固定的频率范围内允许的 FM 发射器的数量要比 AM 发射器少。因此，消耗带宽但质量较高的 FM 调制只在 UKW 频段（超短波）并且是有限的覆盖范围内使用。它是在 20 世纪 50 年代被引入的。

UKW 无线广播的频率从 87.5～108MHz。信息信号的范围是 30Hz～15kHz。为了提高高频信号的抗干扰能力使用了预加重技术，它由时间常数为 50μs 的高通滤波器构成，如图 3-29 所示。偏差为 75kHz，根据式（3-26）FM 信号的带宽为 180kHz。

在 20 世纪 60 年代引入了立体声广播。为此人们寻找到一种方法，使旧的单声道设备仍然能够继续接收广播。

立体声信号由两个声道组成，即 L 信号（左声道）和 R 信号（右声道）。根据这两个信号，首先产生一个 S 信号（和）和一个 D 信号（差），计算方法如式（5-1）。

$$S=\frac{1}{2}(L+R) \qquad R=S-D$$
$$D=\frac{1}{2}(L-R) \qquad L=S+D \tag{5-1}$$

两个信号都要经过频分复用来传输。S 信号占用频率一直到 15kHz 的基带。D 信号将 38kHz 的辅助载波调制到抑制载波的 ZSB-AM（DSSC）上。因此产生范围为 23 ~ 38kHz 及 38 ~ 53kHz 的边带。这个边带不在 S 信号的范围内，因此 S 信号可以与 DSSC 信号相加。

载波被 DSSC 抑制掉，这是发射器的调制范围所希望的。作为载波的替代，在 19kHz 处插入一个导频，接收器可根据它产生原始的载波。

图 5-1 给出了所谓的立体声复用信号的频谱，它由 S 信号、D 信号和导频组成。此时 UKW 发射器使用 FM 进行调制。

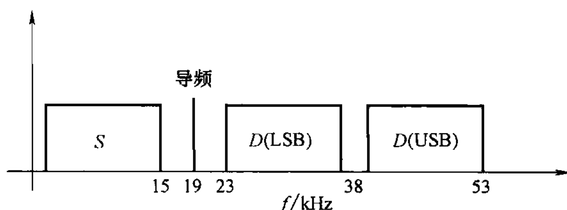


图 5-1 立体声复用信号的组成

因为通常情况下 D 信号的幅度比 S 信号的幅度小很多，在实际中立体声发射器与单声道发射器的传输带宽相等。使用 75kHz 的偏差能够产生 180kHz 的带宽。

接收器对 FM 信号解调后又产生了立体声复用信号。单声道接收器将 S 信号过滤出来，而立体声解码器中的立体声接收器能够产生出 L 信号和 R 信号。

从 20 世纪 70 年代中期开始引入了 ARI 系统（针对汽车司机的无线信息广播）。这些系统允许发射器识别街道交通信息并在相应的时刻对这些信息进行广播。这样就能够控制接收器（从 CD 或自动盒带中切换到 FM，有可能还会提高放大器的音量）。对 ARI 来说，在立体声系统之上又在 57kHz 处增加一个导频。在广播交通信息时这个导频使用 125Hz 进行调幅。在 23.75 ~ 53.98Hz 这个频段内进行额外的幅度调制还能识别区域信息。

RDS（Radio Data System）代表了 FM 的进一步的发展。它能提供节目的标识并且能使用另一个发送频率发送出同一节目的详细信息，如节目的类型（新闻、喜欢的音乐类型等）、时间、差分 GPS 导航系统的参考值等。RDS 利用数字 ZSB 方法来实现，它使用 57kHz、与 ARI 载波相比旋转 90°的载波进行调制。ARI 和 RDS 相结合产生了图 3-21 所示的 QAM。

3. 数字音频广播（DAB）

UKW-FM 系统为家庭用户提供了从目前的角度来看声音质量还勉强过得去的广播业务。但在移动情况下接收会产生一些问题，因这在用户中占很大一部分比例，因此是很重要的。这个问题是由矮小的山头、建筑物等造成的反射而引起的多径接收产生的。由此而引起的干扰使在汽车中的广播接收很难完成。将频率变换到另外一个、瞬时接收效果比较好的发射器中可以减缓这种现象，采用 RDS 方法这可以自动完成。另外一种对接收器来说比较昂贵的解决办法是采用 4.3.6 节中所介绍的空间分集方法。

第二个问题是频率资源比较紧张。几年前 UKW 无线广播的频率上限已经从 104MHz 扩展到 108MHz，但大量建设新的本地广播站并不能缓解频率资源紧张这种现象。

第三个问题是 UKW-FM 系统不支持多媒体。数字附加业务如交通信息等能给移动用户带来很多应用。

很明显，对 UKW-FM 系统的不断演进并不能到达预期的目标。因此应该对它进行革命性的改进，由此而引起的向下不兼容也是必要的。改进后的系统当然应该采用纯数字的方式工作。这种新的系统称为 DAB（Digital Audio Broadcasting），目前已经投入使用。

频率资源紧张的状况要求使用信源编码。DAB 使用 MPEG-2 第二层定义的 MUSICAM 方法，参见 4.1.4 节。对立体声信号来说数据速率在 160～256kBit/s 之间。

原则上对多媒体的支持在数字传输中已经介绍过了。难点在于数据传输速率是不同的并且部分值可能是巨大的。音乐的乐谱及手稿很容易随着音乐一起传输。这些附属于收听节目的数据称为 PAD（Programme Associated Data）。除此之外还可传一些不相关的数据，如交通信息、股票行情、旅游信息等。根据需要可对这些信息进行加密，需要安装附加的设备才能接收加密的数据。与节目无关的数据可采用异步的方式传输。因此可采用一个独立的复用器将各种分组复用在一起。不同的数据对可靠性的要求不同，因此有它们自己的信道编码器。

根据信息内容的不同信道编码有不同的复杂性。可使用采用维特比（Viterbi）解码器的点卷积码。信道编码和时间交织产生的时延约为 0.5s，需要例如通过播音室的时间信号中进行补偿。

多径接收造成的问题可用 COFDM 方法来解决，参见 3.4.6 节。带宽固定在 1.536MHz，可以分成间隔为 1、2、4 或 8kHz 的 1536、768、384 或 192 个子载波。每个子载波采用差分 QPSK 进行调制。实现的数据传输速率超过 2MBit/s，远大于经过压缩的立体声信号所需要的速率。因此人们将 6 个无关的立体声节目和它们的 PAD，以及其他数据封装成一个 DAB 群（DAB-Ensemble）。如图 5-2 所示为数据流的构成方法。

DAB 群的划分是很灵活的，如体育节目的传输速率可比音乐节目的传输速率小。每帧的开始为一个空字符和一个发送暂停，这能使接收器对当前的干扰情况进行判断。接下来就是正弦波符号，接收器可以据此来确定它们的均衡器。再后面就是快速信息信道（FIC），它包含复用帧及当前使用的信道编码。

图 5-2 所示的 CIF 数据会交给 OFDM 调制。由于有保护间隔（参见 3.4.6 节），OFDM 信号对多径接收不敏感。这允许构建一个单频网络（Single Frequency Network，SFN）。因此同一个 DAB 群可以在最远相距 70km 的发射器中很精确地进行同步并在同一个频率上发送出去。这种方式是除了信源编码之外又一种节省频率资源的方法。除此之外，与 FM 相比 DAB 的发送功率也可以降低，因为除了编码增益之外还有网络增益，多个发射器发出的信号可以相互支持。

目前使用陆地单频网络传输的 DAB 节目被频率约为 220MHz 的 TV 频道 12 接收，它包括 DAB 群，确切的频率分配目前还是国际上争论的热点。目前计划通过卫星来传

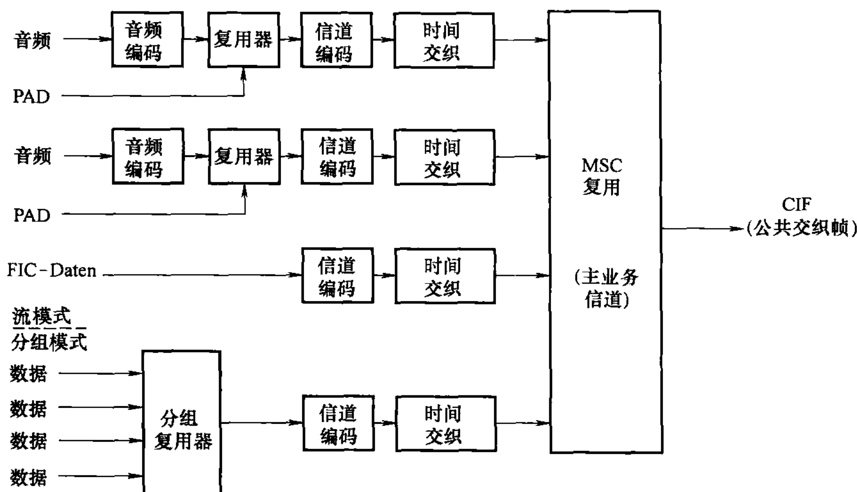


图 5-2 DAB 信号复用的结构图（总共有 6 个音频/PAD 信道，这里只画了两个）

输 DAB，采用频率为 1.5GHz（L 波段，比较表 1-11），带宽为 40MHz 的宽带信道。

从商业的角度来看，DAB 并没有取得所希望的那么大的成功。一方面是由于覆盖范围比较小，且接收器价格较高，另一方面，坦率地说，数据部分（分组模式）还缺乏有趣的内容。从技术的角度上来说，DAB 也需要批评：与 DVB-T 太相近（参见 5.1.2 节），最终电视会覆盖广播。

4. 数字长波、中波、短波广播

从表 5-1 可以看出，AM 的数字继任者要使用一个很难使用的信道。可能的多径接收会引起码间干扰（ISI），为此需要使用均衡器。信道的时变特性甚至要求使用自适应的均衡器。只有使用信源编码，将信息信号大幅度压缩，才能将数据的传输保持在很窄的带宽上。由于受很强的大气空间的干扰信号的影响，所以这种方法要求使用信道编码。

但系统的设计并不仅基于技术条件，而是面向市场。在人口稠密、高度工业化的国家需要通过电缆、VHF 无线广播及卫星来实现广播的覆盖。在人口稀疏的地区，由于铺设电缆网络、VHF/UHF 网络等受覆盖面积的影响投资是很高的，因此还使用卫星、长距离的长波、中波及短波进行发送。使用卫星发送可以享有高带宽、没有变化及干扰小的信道。因此只有当接收器确实比卫星接收设备便宜许多时，30MHz 以下的数字广播才有机会。可以很明确地说：30MHz 以下数字广播的主要接收者是在发展中国家。接收器的价格必须要在 10 ~ 20 欧元之间，因为在发展中国家几乎没有人愿意将一个月的收入投入到一个接收器中。而且接收器的工作还要比较省电，因为电池等短缺商品，且电网并不能到达每个人。

由于这些原因，DAB 早已被引入到市场中而 30MHz 以下的数字广播在 2001 年才被标准化。不过至少在这点上人们愿意抓住机会，在整个世界范围内采用统一的标准。

2003 年新的系统投入运营。

在长波、中波及短波波段上不能像引入 DAB 那样简单地使用新的频率。在选择新的系统时如何进行过渡是很关键的。

在 ITU-R 下属的一个机构 DRM（Digital Radio Mondial）联盟中，共给出了 5 种相互竞争的建议。其中的 3 种基于多载波方法，即 COFDM，另外 2 种倡导单载波方法，即多元 QAM。后面这组需要复杂的均衡器，而前面一组可以节省传输符号持续时间。前一组中的 Skywave2000 系统在竞争中取胜：

1）音频数据首先根据 MPEG-2 第 3 层进行压缩。

2）接下来的信道编码包括卷积编码器和交织器，接收器采用 Viterbi 算法进行解码。

3）根据一组载波构造出 HF 信号。一个所谓的基本单元由间距为 33.3Hz 的 96 个子载波组成的核心构成，相当于 30ms 的传输符号持续时间及 3200Hz 的带宽。这个核心可以灵活地添加 11 个子载波构成的组。例如使用 16 个附加组可以构成总共有 272 个子载波、总带宽为 9kHz 的信道，这适用于今天的 AM 栅格。

子载波使用 16QAM、64QAM 或 256QAM 对音频数据进行调制。256QAM 是为地面波传输扩展而考虑的。与 DAB 相同，这里也携带 PAD 信息。

4）核心的 3 个子载波不进行调制，它们作为消除多普勒效应的参考频率使用。

5.1.2 电视

1. 黑白电视

在美国、英格兰、法国和前苏联，从 1948 年开始就有了电视广播。瑞士从 1953 年开始才有了有规律的电视广播。这里的“有规律”是指每周三个晚上，每个晚上 1 小时的节目，当然是黑白的。直到 1960 年视频信号还不能用磁带存储，除故事片之外所有的内容都是纯粹的直播。从 1960 年开始有了存储产品，节目广播发生了翻天覆地的变化。1968 年开始有了彩色电视。

视频信号是二维的，但为了传输要将它转变成一维信号。在摄像机中对图像按行进行采样，各行按序组成串。为了重构图像，要插入同步信息（水平即行同步及垂直即图像同步），如图 5-3 所示。

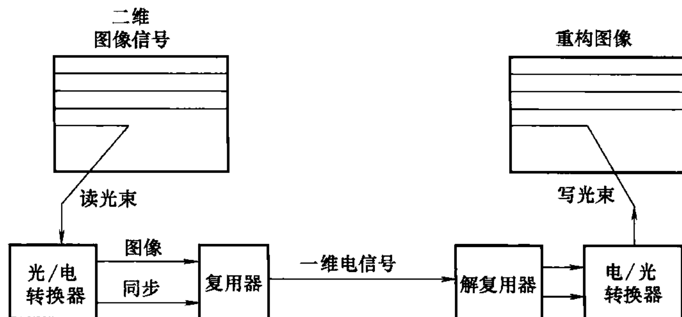


图 5-3　图像传输的原理

在欧洲每幅图像按625行进行采样，它确定了可达到的图像解析率。所有调整到同一个节目上的接收器，它的显像管的电子束都与摄像机在图像平面上扫描的电子束是同步的。电子束的位置由两个锯齿产生器来确定，而锯齿产生器由同步信号来控制。每个图像点的余辉持续时间必须与图像持续时间严格匹配，这样发光可以持续一点点时间但又不至于在图像移动时在边缘产生阴影。

每秒钟必须发送50幅图像，这样人的眼睛才不会感觉到光的闪动。但这要求很大的传输带宽。因此作为一种折中引入了跳行的方法，即每幅图像被分成由偶数行和奇数行组成的两个半幅图像。

这样图像扫描的频率变成25Hz，垂直频率为50Hz，水平频率为625 × 25Hz = 15625Hz。传输一行的时间包括同步在内为64μs。

接收器的中间存储及重读同一图像可以减弱闪烁。带有数字中间存储的现代接收器也采用这个办法。电影院的放映机每秒钟播放24幅图像，使用所谓的无闪烁技术放映机在播放每幅图像时都进行短暂的停顿，这种方法产生的效果与双倍图像频率产生的效果基本相同。

黑白电视中图像信号使用所谓的BAS信号（图像、消隐间隔、同步）来描述。图像点的亮度由电压大小的级别来确定，即0%（黑），100%（白）。然后是电压为0%消隐间隔，这样在换行时写射束是黑的。行同步的电压级别为-43%（比黑还黑）。同步脉冲之后是作为参考电压的压级为0%的后沿。垂直同步通过重传同步脉冲（压鼓替代击鼓）。图5-4所示为行同步的过程，它占用为每行预留的64μs中的12μs。图中所示的突发脉冲是彩色图像调制时所需要的，我们将在后面讨论。在图5-4所示的同步阶段之前是第 n 行，然后是第 $n+2$ 行（跳行方法）。

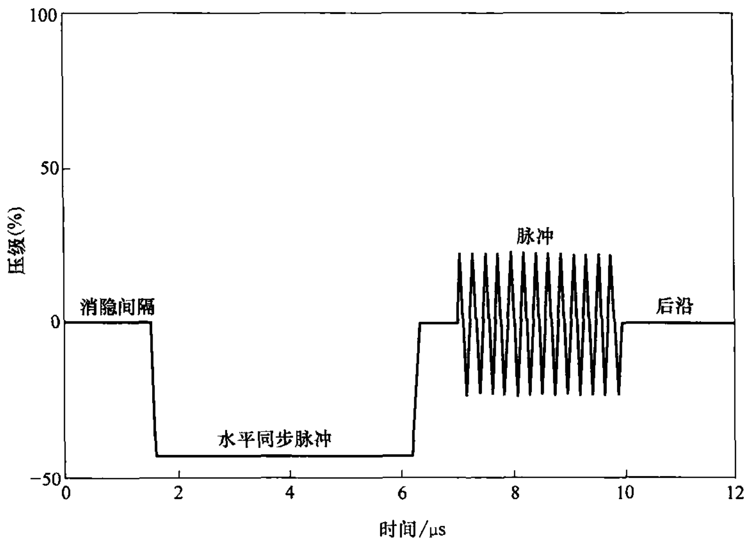


图5-4 行同步

BAS 信号的带宽是 5MHz。这个数字不包括同步带来的复杂性，是按照下面的方法估计的：625 行决定了垂直方向的解析率。在相同的水平解析率及 4:3 的宽高比情况下，每行有 625 × (4/3) ≈ 833 个图像点，这样每幅图像共有 625 × 833 = 520625[⊖]个点。当每秒传 25 帧时每秒钟共产生 1300 万个点需要传输。如采用 PAM 对这些点进行模拟传输，最少需要 6.5MHz 的带宽。但这只是当每个点轮流是亮的或暗的时才是必需的。因此在没有较大质量损失的情况下带宽限制在 5MHz。

在无线传输及经过频分复用的有线传输的情况下，需要对 BAS 信号进行调制。由于 BAS 信号的带宽很大，因此不考虑使用 FM。使用 AM 也是不经济的，因为它需要双倍的带宽。BAS 信号可能包含低频成分，所以 SSB 是不能实现的。因此只剩下 VSB（残留边带，参见 3.1.3 节）。使用这种调制方法带宽增加到 7MHz，载波会随之传输。因为接收器必须使用乘法解调器，因此需要相位很准的载波。

BAS 信号在调制之前要进行翻转，即高幅度的 HF 对应于比较暗的点。根据调制深度可以确定下面的幅度：同步 = 100%，后沿 = 75%，白 = 10%。同步时周期性的峰值可以给接收器一个用于调整它的接收强度的判决标准，这种方法可以对与距离有关的传输衰减进行补偿。

声音信号采用 FM 图像载波之上的 5.5MHz 的声音载波进行传输，所占带宽为 50kHz，如图 5-5 所示。

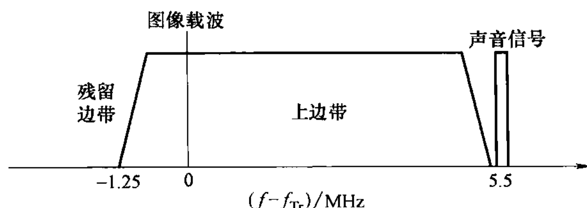


图 5-5　带有声音的已调 TV 信号（黑/白）的频谱

2. 彩色电视

彩色图像传输时每幅图像被分成红、绿、蓝 3 个部分。由此产生 3 个信道的 RGB 信号，接收器根据这个信号合成原始的彩色图像。彩色显像管由 3 个并行的系统组成，每个系统有一个电子束来负责这 3 种颜色。

像 FM 立体声传输一样，这里的信号也要进行转换。RGB 信号本身由于下面两个原因不适合传输：

1）3 个信道中的每一个带宽都为 5MHz，是黑白电视传输带宽的 3 倍。在 4.1.5 节中已经介绍过了，这是没有必要的。

2）黑白电视接收机不能处理 RGB 信号。

因此 RGB 信号要根据下面的公式转换成亮度信号 Y（亮度信息）：

$$Y = 0.3R + 0.59G + 0.11B \tag{5-2}$$

式（5-2）中的系数是由人眼对颜色的敏感度得来的。这个 Y 信号保证了对黑白接收器的兼容性。

⊖ 原书为 625 × 833 = 520000，有误。——译者注

然后可以导出颜色差信号 U 和 V：

$$U = B - Y \qquad V = R - Y \tag{5-3}$$

接收器可以根据 Y、U、V 重新计算出 R、G、B。

发射器使用低通滤波器将信号 U、V 的带宽减小到 1.3MHz，因为如 4.1.2 节所述，人眼对颜色的分辨功能相对较差。

接下来两个频率为 4429687.5Hz（简化为 4.43MHz）的正交载波与 U 和 V 信号进行 DSSC 调制。这是如图 3.21 所示的正交 AM（QAM）。这个 QAM 信号称为色度信号，包含颜色信息。

现在将亮度信号（范围为 0～5MHz）和色度信号（范围为 4.43MHz ± 1.3MHz）叠加到一起形成 FBAS 信号（彩色 BAS）。两个信号的频谱相互交织在一起，乍看起来要在接收端进行分离似乎不太可能。但实际上由于图像是按行构成的，信号 Y、U、V 实质是周期信号，周期同 64μs 的行持续时间相等。它是具有线间隔 1/(64μs) = 15.625 kHz 的线频谱。色度信号也是如此。它的载频正好是 283.5 × 15625Hz，因此色度信号的频谱线正好在亮度信号的频谱线之间。

FBAS 信号采用跟前面所说的图像一样的 VSB 调制方法，FM 声音信号在 5.5MHz 处加到图像载波上。图 5-6 给出了它的频谱情况。

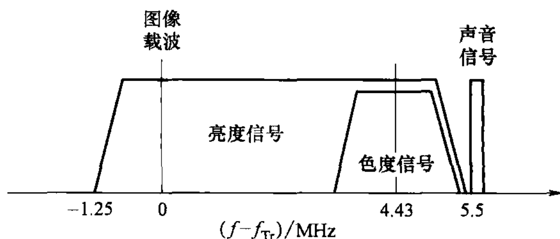


图 5-6　彩色电视信号的频谱

接收器对 VSB 信号进行解调，重新获得 FBAS 信号以及 FM 调制的声音信号。使用 QAM 解调器可以产生出 U 信号和 V 信号。这个解调器必须与发射器中的 4.43MHz 的振荡器的相位严格匹配。因此为了解调器的同步在行同步信号中需要传输所谓的突发脉冲，如图 5-4 所示。

解调后色度信号的相位误差会导致所显示的图像的颜色出现错误。以前按照美国 NTSC（National Television System Committee）标准生产的设备都有一个调整颜色的按钮。目前 NTSC 发射器在垂直消隐间隔会传输一个参考信号。

德国的 PAL（Phase Alternating Line）系统发展了 NTSC 方法。4.43MHz 的载频从一个图像行到另一个图像行进行 180°的翻转。接收器对相位跳变进行跟踪，当前的相位由突发脉冲来通知。如果传输中出现相位错误，颜色从行到行逆向变化，颜色的误差发生在中间部分。

625 行并没有都真正显示在屏幕上。在两帧之间的空隙可以传输一些附加信息，这些信息在德国称之为 Videotex，在瑞士称之为 Teletext。在两个半幅图像之间，发完第 7

行后，发第 8 行数据之前传输这些文字信息。还有一行要留给 VPS（Video Programming System）使用。还有两行留给用于线路控制的检测信号及调整参考来使用，如反射和噪声测量。

宽带电视信号的传输只能考虑使用 VHF 频段的频率。表 5-2 所示是电视使用的频带。

表 5-2 电视频带的划分

范围		频道	栅格/MHz	频率/MHz
Ⅰ	VHF	2 ~ 4	7	47 ~ 68
Ⅲ	VHF	5 ~ 12	7	174 ~ 223
Ⅳ	UHF	21 ~ 34	8	470 ~ 582
Ⅴ	UHF	35 ~ 69	8	582 ~ 862

关于表 5-2 还有下面的补充说明：

1）无线广播频带Ⅱ属于 UKW-FM 范围，频率是 87.5 ~ 108MHz。

2）下面的频道是被禁止的：1（太窄）、12（目前给 DAB 保留，参见 5.1.1 节）、36（给雷达预留）、38（给射电天文学预留）。

3）该表格适用于无线地面广播。有线电缆网络使用其他频率（特殊频道），卫星直接接收使用 12GHz，它有很大的天线增益。

可以保证的最小接收信号强度是在 65dBμV/m（频段Ⅳ）及 70dBμV/m（频段Ⅴ），它们是距地面 10m 高度的测量值。电视转换器根据所处地区的信号覆盖情况功率为 1 ~ 250W。

作为改进引入了 PAL-Plus，它的 16:9 的图像格式比 PAL 系统的 4:3 的图像格式更适合表现人的脸部区域。

更进一步的改进是 HDTV，它主要是为了提高清晰度。基于从未实现的 MAC（Multiplexed Analogue Components）标准开发出 D-MAC，它包括多个数字音频信道和一个模拟图像。HD-MAC 对图像的清晰度又做了进一步的改进，它将行数提高到 1250。

3. 数字电视（DVB）

DVB 意味着完全使用数字的方法在 PCM 中描述信号，这在一些播音室中已经实现。图像信号在传输中使用 4:2:2 的格式来描述（参见 4.1.6 节），数据传输速率为 216MBit/s。使用 MPEG-2 算法可以将速率减少到 9MBit/s（播音室质量）或 6MBit/s（终端用户质量）。之后是一个扰码器，它将信号的能量平分在整个频谱上，最后是一个使用 Reed-Solomon 码的信道编码器和交织器，如图 5-7 所示。

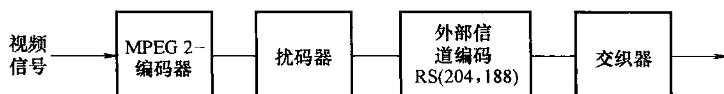


图 5-7 DVB 传输之前信号的预处理

DVB 信号的传输有 3 种方法：卫星（DVB-S）、有线网络（DVB-C）、地面 VHF/UHF 网络（DVB-T）。

每种信道都有它们自己的特点，因此信号按不同的传输方式进行适配。图 5-7 所示的预处理对所有的情况是相同的。

卫星传输（DVB-S）

卫星信道至少在下行链路，即从卫星到地球这段上功率是受限的，这是因为要受太阳能供电系统复杂性的影响。但另一方面，有很大的带宽可供使用，因为从 10.7～12.75GHz 的频段是为卫星直接接收而预留的。并且由于卫星天线有很强的方向性，这个频段可以多次重复使用，即采取空间和极性复用的方式。卫星中的功率问题使用非线性发送级别，可以达到较好的有效系数。但它不允许使用 QAM，在 DVB-S 中使用 QPSK。图 5-8 是 DVB-S 传输的结构图。总的来说在路径上有两个信道编码器，如图 4-24所示。DVB-S 已经投入实际运行。为了能够继续使用传统的电视接收器，由机顶盒来完成这些功能。

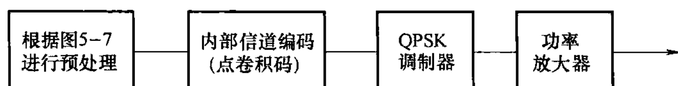


图 5-8 DVB-S 传输的结构图

同轴电缆传输（DVB-C）

由于存在相对较大的衰减，同轴电缆网络包含许多放大器。这些放大器的中间调制产物的值在模拟传输时是很大的，而在数字传输时会减少 10dB，因此中间调制变得不是很重要。在电缆网络中的反射会导致线性失真，即回声。

电缆信道的带宽是有限的，因此需要选择多元的调制方法，即 64QAM。信号与反射信号的电平差最小应该是 30dB，这样反射信号才不起作用。因为误码率相对较小，可以在内部信道编码器（inneren Kanalcoder）中被消除，在调制之前还使用一个差分预编码。这会使接收器变得简单。在 8MHz 宽的信道中有用数据的传输速率可以达到 38MBit/s，因此一套模拟节目可传 4～5 套数字节目，当然这些节目首先必须使用 MPEG-2 MP@ ML 进行压缩。图 5-9 是 DVB-C 传输的结构图。

地面 VHF/UHF 网络传输（DVB-T）

这是一种比较困难的传输方式，因为山坡、建筑物等的反射所引起的多径接收会影响数字信号的接收（ISI）。因此这部分 DVB 系统在最后才被定义，并且尽可能多地借鉴其他标准。DVB-T 应该占用 8MHz 的信道栅格，因此不能使用 VHF 频段 Ⅰ 和 Ⅲ。它的主要用户是装有屋顶天线的家庭接收器（天线增益约为 15dB）。便携式接收器（位置固定，但只是辅助天线或设备天线）只有在发射器附近时才可行。移动接收器不适于在这种情况下使用。这种方法可以使用单频

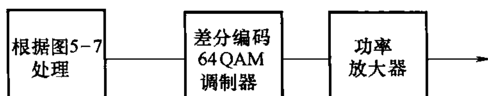


图 5-9 DVB-C 传输的结构图

网络，也可以使用如图 3-74 所示的分级调制来实现。

对单频网络的这种要求自然导致要使用 COFDM。相距 60km 的发射器要求保护间隔约为 200μs，相对较长的符号持续时间导致需要无数的子载波，因而会使设备复杂化。基于这种考虑该标准允许不同的符号持续时间，由每个网络来指定它的系统及调整它的覆盖范围。图 5-10 所示为发送端的信号处理过程。

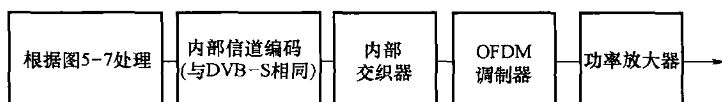


图 5-10　DVB-T 传输的结构图

子载波的数量约为 1700 或 6800，可以采用 QPSK、16QAM、64QAM、MR-16QAM 或 MR-64QAM 进行调制。

发射器处于较高的位置上，而播音室通常位于拥挤的城市中心。因此图 5-10 所示的一系列单元会在内部交织器和 OFDM 调制器处分离，并在这里进行远程传输。可以使用光缆进行点到点的传输，或使用像 ATM（参见 6.3.5 节）那样的数据网络传输。

需要的接收信号强度根据选择的信道编码方式的不同在 31 ~ 55dBμV/m 之间变化。DVB-T 比模拟传输需要的接收信号强度要小得多，但能提供 6 倍的节目数量。DVB 的工作很接近香农极限，但系统很复杂。

在美国 DVB 并没有得到承认。DVB 与 DAB 一样是在欧洲发展起来的，它并没有预见到 HDTV 传输。美国都很重视高清晰度电视，并发展起来数字的基于 8-VSB 调制的 ATSC（Advanced Television Systems Committee）。

尽管某种程度上 DVB-T 标准会使 HDTV 成为可能，但两者可能会同时存在。两者在地理区域上的分布很可能跟今天的 PAL 和 NTSC 的划分相对应。

另外一个有争议的问题是是否还需要继续使用跳行的方法。计算机业界希望将它废除，而无线广播业界考虑到兼容的原因希望将它保留。

5.2　接收技术

5.2.1　直接接收器、单变频及双变频超外差接收器

直接接收器是形式最简单的接收器，但它已经不能满足当今的需求。尽管如此，作为讲解还是值得简单地将它介绍一下。图 5-11 是它的结构图。

天线接收的信号首先要经过带通滤波器，它完成两个功能：选择希望的发射器（信道分离）、改善信噪比（比较图 1-67）。

带通滤波器之后是一个 HF 放大器，接着是一个解调器，最后是一个 NF 放大器。这里还有一个反馈环，它能改善输入范围内的衰减、质量及由此产生的信号分离的程度（邻近选择性、受邻居信号的压抑性）和敏感性。使用这个原理的接收器被称为晶

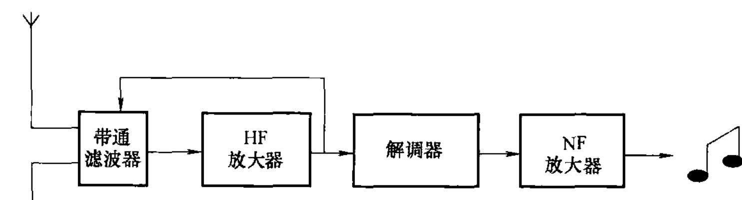


图 5-11 直接接收器

体管接收器，但由于调节是非常困难的，它不能强调反馈环。如果反馈作用太小，接收器会很不敏感。如果反馈太大，接收电路会开始振荡，接收器会变成发送器。

当天线接收到的信号的功率很大时，甚至可将图 5-11 所示的两个放大器都省略掉，变成纯无源接收器。图 5-12 所示的中波（MW）接收器电路现在的中小学生还在搭建它。解调器使用图 3-13 所示的检波器，考虑到可携带性，耳机膜无需使用特殊的滤波。经过解调后 DC 分量也到达耳机，但这并不妨碍任何事情。用一个锗或肖特基二极管比较好，它们的正向电压降比硅二极管的电压低。耳机必须是高阻抗（2kΩ），振荡电路的衰减要尽可能地小。

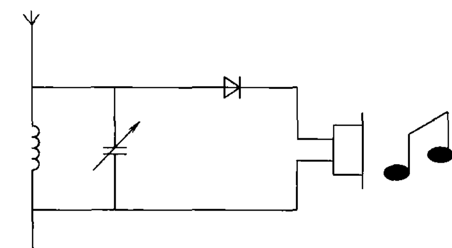


图 5-12 最简单的中波接收器

仅仅是一个趣事：最简单的 AM 接收器可以由一个新鲜的柠檬构成，柠檬里面插着一个铜线和一把厨房用的刀。耳机就连接到这两个金属上。这是一个检波器，因为两种不同的金属放在酸性环境中，会产生一种非线性电化学元素。当然这种原理只工作在离 AM 发射器很近的环境中。

直接接收器的弱点在它的输入滤波器上。它必须为每个发射器进行相应的设置，因此它的质量及其振荡电路的相对带宽才能保持基本恒定。但希望的是能够有一个绝对恒定的带宽，它与所采用的调制方式相对应。此外当频率较高时，信号的严格分离根本达不到。同时总的 HF 放大集中到同一个频率上。不希望发生的接收器内部的反馈引起放大器振荡的危险很大。超外差式接收器能够解决这两个问题，它能对信号进行很好的分离并且能提供较高的敏感性。

超外差式接收器的技巧在于解调器可以工作在一个固定的频率上，即所谓的中间频率（Intermediate Frequency，IF）。希望的发送器的信号通过两个混频器转变成两个频率，如图 5-13 所示。

图 5-13 中从 ZF 滤波器的输入端开始就与图 5-11 所示的结构图一模一样。之前的各模块的作用是将频率从 HF 转换到 ZF。通过调整本地振荡器的频率可以选择发射器。因为混频器是不清晰的，所以之前必须有一个 HF 滤波器（输入滤波器）。但这不要求一定是窄带的。超外差式接收器会将 ZF 滤波器之间的 HF 频谱覆盖掉，而直接接收器的输入滤波器接收 HF 频谱。表 5-3 给出一些 MW 接收器频率的例子，它们分别位于图

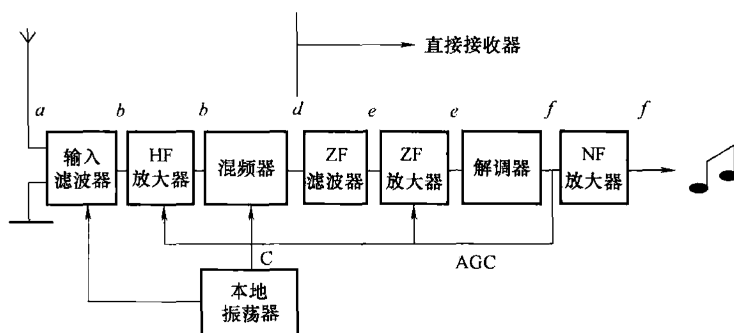


图 5-13　单变频超外差接收器

5-13 中用斜体字表示的各点上。

表 5-3　图 5-13 中的频率

图 5-13 中的各点	涉及的频率/kHz	图 5-13 中的各点	涉及的频率/kHz
a	所有	*d*	和及差频
b	500 ~ 1650	*e*	450 ~ 460
c	925 ~ 2055	*f*	0 ~ 4.5

解调器之后会得到一个 AGC（Automatic Gain Control）信号，用它来调节 HF 和 ZF 放大器。根据发送功率、距离及扩展衰减的不同，不同发射器所得到的调节程度差能达到 100dB。ZF 放大器必须是一个具有很大调节范围的窄带放大器，这样解调器才能一直工作在强度大致相等的信号环境下。ZF 放大器通常采用多级的方式来实现。

超外差式接收器一个最大的缺点就是镜像频率接收（英文为 image frequency），因为在混频时会不断产生和频和差频，因此会有两种可能来接收到发射器的内容。例如一个发射器的频率为 525kHz，一个接收器使用的 ZF 为 455kHz：

差频的构成：　　LO 频率 = 525kHz - 455kHz = 70kHz

和频的构成：　　LO 频率 = 525kHz + 455kHz = 980kHz

在实际中优先选择差频，因为 LO 必须被覆盖掉的百分比很小，在技术上比较容易实现。另一方面，两个发射器可能同时产生相同的频率。当 LO 的频率为 980kHz 时，会有：

希望的发射器：　525kHz→ZF = 980kHz - 5255kHz = 455kHz

镜像频率：　　1345kHz→ZF = 1435kHz - 980kHz = 455kHz

消除镜像频率接收的任务要由输入滤波器来完成。上面计算出的两个频率都处在 MW 频段，滤波器还必须根据所希望接收的频率来进行转换，这点在图 5-13 中也画出来了。

不用对滤波器进行调节的另外一种方法是基于这样的认识：希望的频率与镜像频率之差刚好是中间频率的两倍，如图 5-14 所示。只要将中间频率设得很高，输入端接

收滤波器就可以不用进行转换。

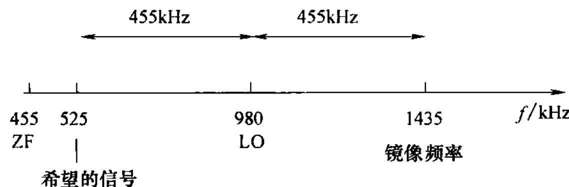


图 5-14 镜像频率的位置

因此中间频率的选择是对下面两个方面的一个折中：

1）ZF 尽可能高→简单的输入接收器，对镜像频率有较好的抑制效果。

2）ZF 尽可能低→ZF 滤波器简单，邻近选择性好。

将“频率转换”和“滤波”的任务划分为两块在图 3-16 的 SSB 调制器中我们已经遇到过了。

最好的，但也是最贵的解决 ZF 选择问题的方法是使用双变频超外差接收器。它使用两级超外差接收器，将第一级 ZF 设置得比较高（较好的抑制镜像频率的效果），第二级 ZF 设置得较低（较好的分离效果）。图 5-15 为它的结构图。如果使用乘法解调器的话，在信号通路上就会共有 3 个混频器。

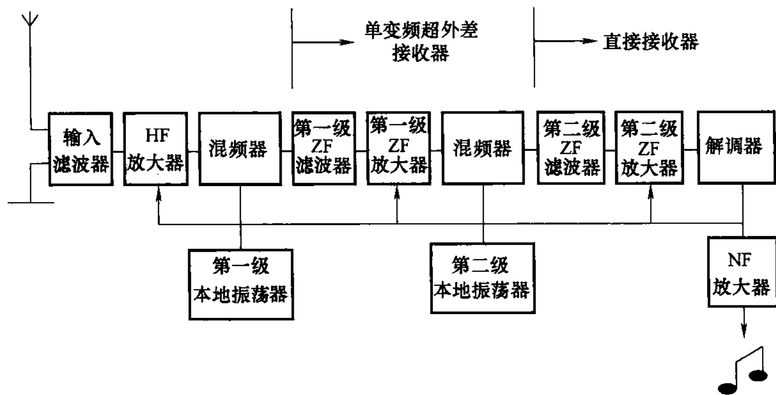


图 5-15 双变频超外差接收器的结构图

为了设置希望的发射器，两个本地振荡器中的一个的频率必须是可变的。目前优先选择第一级本地振荡器可变，这样后面各级不用接收过多的信号。目前的振荡器使用基于 DDS 原理的合成方法来实现，参见 5.4 节。

不同的 ZF 级别必须很好地区别开来。如果一个信号直接加到某一级的 ZF 上，那么与振荡器的设置无关，在扬声器中一直都能听到这个信号。因此在接收器的输入端经常根据规模设置 ZF 所确定的吸收电路。表 5-4 给出了几个经常使用的中间频率。

接收器安装的主要难点与频率范围有关：

长波、中波接收器

这些接收器很容易实现，因为它们主要用于无线广播，因此能够在固定的信道栅格上处理很强的信号。由于大气的干扰噪声在这个频段内很强，接收器不需要小噪声电路。

表 5-4　经常使用的中间频率

中间频率	应　用
455kHz	MW 单变频超外差接收器，双变频超外差接收器的第二级 ZF
9MHz	UKW 接收器
10.7MHz	专业接收器的第二级 ZF
38.9MHz	TV 接收器
45MHz	KW 双变频超外差接收器的第二级 ZF
70MHz	专业接收器的第一级 ZF 定向广播设备（微波接收）

短波接收器

在短波波段内实现接收器是很重要的，大气干扰噪声与半导体和电阻噪声相比仍然很大。在这点上没有什么挑战。主要问题还是不同发射器在相邻频率上的功率差特别大。如果将接收器设置在一个信号比较弱的发射器上，AGC 会将 HF 和 ZF 放大器调节得很大。相邻的信号比较强的发射器的 HF 放大器和第一级混频器就会出现过调节，由此而引起的中间调制产物就会将有用的但较弱的信号覆盖（“堵塞”了接收器）。信号大小的一致性是高质量 KW 接收器质量评价的最重要的标准，它可以通过下面的几种方法得到改善：

1） 仔细设计具有高动态范围的电路。

2） 早期滤波，也就是说，在宽带接收器中输入滤波器是可以开、关的。

3） 可变换的衰减等级。这些方法在 1.2.3 节第 3 部分的结尾部分已经描述过了。

图 5-16 所示为现代短波接收器的输入部分。

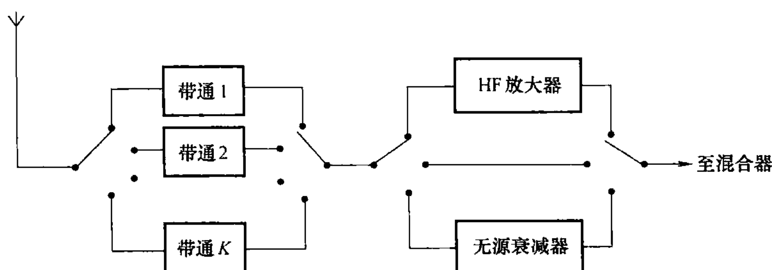


图 5-16　使用可调节输入带通滤波器及可选择 HF 放大器和衰减器的现代短波接收器的输入部分

KW 接收器的另一个难点是频带重叠得很厉害，它最多只允许无干扰地接收无线广播。通常可以使用复杂的 ZF 滤波器来解决这些干扰，值得一提的是：

1） Notch 滤波器（抑制滤波器，即窄带、可调节的带阻，以消除在 AM 或 SSB 接收时由很难听的尖叫声所形成的陌生载波）。

2） 噪声消隐器（消除脉冲类型干扰的消隐器，放在窄带滤波器之前，这样可以缩短反应时间）。

3）边缘可调的 ZF 滤波器（调节可以通过的带宽）。

也可以在 NF 范围的解调之后采用比较便宜的方式实现滤波功能。但 ZF 方法有较大的优势，因为此时 AGC 是真正希望的信号。顺便说一下，使用 NF 滤波器的、比较便宜的接收器已经有了改进，目前已经开始使用数字信号处理（DSP）方法来实现它。

VHF 接收器

当频率增高时大气干扰噪声的强度会下降。因此对接收器随频率变化的敏感性的要求增大。在 VHF 频段内有许多无线广播业务，因此大多数情况下信号都很强，并且相对容易接收。特殊情况下直接在天线处增加一个低噪声的前置放大器（Low Noise Amplifier，LNA）。它的任务是对频率增大所产生的集肤效应以及逐渐增大的天线到线路的损耗进行补偿。如果将放大器放在接收器的输入端，虽然信号的强度会有所增加，但电缆衰减造成的信噪比会变差。

UHF 和 SHF 接收器

这些接收器的关键指标是由接收器的固有噪声所决定的敏感性。接收器固有噪声的主要部分是由 HF 放大器所引起的，因此在有特殊要求的情况下（例如射电天文学）要将它们进行冷却，参见式（1-20）。

一个简单接收器的标准指标包括例如接收信噪比为 10dB 的 SSB 信号的敏感度为 0.25μV。在通常波阻抗为 50Ω 的情况下所需要的信号功率为

$$P_S = \frac{U_{eff}^2}{R} = \frac{(0.25\times10^{-6})^2}{50}\mathrm{W} = 1.25\times10^{-15}\mathrm{W} \triangleq -119\mathrm{dBm} \tag{5-4}$$

因为 SSB 没有任何调制增益，接收器的噪声 P_N 要求比指标低 10dB，因此值为 $1.25\times10^{-16}\mathrm{W}\approx-129\mathrm{dBm}$。

接收器噪声的理论极限值 $P_{N_{min}}$ 由式（1-20）给出，当温度为 290K（绝对温标）、带宽为 2.7kHz 时：

$$\begin{aligned} P_{N_{min}} &= kTB = N_0B = 1.38\times10^{-23}\times290\times2700\mathrm{W} \\ &= 4\times10^{-21}\times2700\mathrm{W} = 10.8\times10^{-18}\mathrm{W} \triangleq -139.7\mathrm{dBm} \end{aligned} \tag{5-5}$$

（注：表示温度时人们喜欢用 290K，而不是通常的温度 293K（20℃），因此 N_0 的值约为 $4\times10^{-21}\mathrm{W/Hz}$）。

接收器的固有噪声将信噪比减少了大约 10dB。实际的（即有噪声的）接收器是用没有噪声的接收器并在输入端口外加一个噪声源来进行建模的，后者具有噪声功率谱密度 N_E 及噪声功率 $P_E = N_EB$。总的接收噪声为

$$P_N = P_{N_{min}} + P_E = N_0B + N_EB$$

现在使用理论最小值的 n 倍来表示总噪声：

$$P_N = nP_{N_{min}} = nN_0B = N_0B + N_EB$$

$$n = \frac{N_0 + N_E}{N_0} = 1 + \frac{N_E}{N_0} = 1 + \frac{P_E}{P_{N_{min}}} = \frac{P_N}{P_{N_{min}}} \tag{5-6}$$

$$F = 10\lg(n) \quad \mathrm{dB} \tag{5-7}$$

其中，n 称为噪声系数，F 称为噪声值。对我们上面的接收器来说，有

$$n=\frac{1.25\times10^{-16}}{10.8\times10^{-18}}=11.57 \quad 和 \quad F=10\lg(n)=10.6\text{dB}$$

噪声温度 T_R 是一个概念温度，它用来控制噪声源 N_E（作为热噪声源来观察），因此还会有一个附加的热噪声功率：

$$N_E B = P_N - N_0 B = nN_0 B - N_0 B = (n-1)N_0 B$$
$$N_E = (n-1)N_0$$
$$kT_R = (n-1)kT \Rightarrow T_R = (n-1)T$$

上面接收器的噪声温度值为（11.57 - 1）×290≈3065K。n、F 和 T_R 属于不同的"货币"，但描述同一个现象，即接收器的固有噪声。T_R 的分辨率最大，F 通常变化较小。根据式（5-6）用 dBm 表示的一个接收器的噪声集合（总噪声）为

$$P_N[\text{dBm}] = 10\lg(nP_{N_{min}}[\text{mW}]) = 10\lg(nN_0[\text{mW/Hz}]B[\text{Hz}])$$
$$= 10\lg(N_0[\text{mW/Hz}]) + 10\lg(B[\text{Hz}]) + 10\lg(n)$$
$$P_N[\text{dBm}] = -174\text{dBm} + 10\lg(B[\text{Hz}]) + F[\text{dB}] \quad (5\text{-}8)$$

另外还要特别注意振荡器的实现，因为它需要较大的相对频率准确性和稳定性。5.4 节将重点强调这个问题。如果本地振荡器给相互重叠的接收器提供多个频率，邻居发送器会与 ZF 相混淆，信噪比会变差（因此会增大接收器的噪声值）。

不论如何，中波接收器在天线处都需要一个 LNA。在这个频段上信号在同轴电缆的损耗高得令人无法忍受。这个前置放大器也有一个固有噪声，它也能使接收信号的信噪比变差。这个固有噪声是通过信噪比的下降来定义的：F = 0dB 时的 LNA 是无噪声的，F = 3dB 能将接收信号的信噪比降低 3dB。好的前置放大器的 F 会小于 0.5dB，多数情况下使用 GaAs FET 或作为参变量放大器来实现。

从 3GHz 开始可以使用波导管。这种比较昂贵的解决方案只有将发送设备和接收设备结合起来才是值得的。对纯接收系统而言，比较好的方法是用 LNB（Low Noise Block）或 LNC（Low Noise Converter）来代替 LNA。它们将放大器和混频器结合起来，将接收到的信号转换到较低的频率上去，这样在同轴电缆上的损耗就不起多大的作用。原则上将图 5-15 所示的超外差接收器在第一级 ZF 进行剥离并在空间上分隔开即可。

5.2.2 数字接收器

在接收技术上也不能阻止数字化的进程。多年前就已经能够使用数字化的方法产生所有必需的频率，早些时候是使用 PLL（Phase Locked Loop），现在是使用 DDS（Direct Digital Synthesis），参见第 5.4 节。因此也可以毫无问题地通过数字接口来远程控制接收频率。这点是很重要的，因为在短波波段工作频率需要经常改变，频率选择及连接建立越来越自动化。建立起来的还包括已经提到的数字滤波器，它连接到一个接收器的 NF 输出端，作用是消除干扰。

到完全实现数字接收器还需要一些时间，主要限制是 A/D 转换器的采样频率太高、所需的字宽（动态的）太大及其对处理器的计算能力要求太高。

已经能够做到的是使用传统的 HF 和 ZF，在第二级 ZF 滤波器之后对信号进行数字

化。因为它是一个带通信号，所以在滤波器之后立即对它进行所讨论过的欠采样是可行的。ZF 放大器的 AGC 减小了后面数字部分所必需的动态性。特别有优势的是 3.1.6 节所介绍的正交描述方法，因为使用它可以直接将它混频到基带。这相当于一个中间频率为零的超外差接收器，直接混频器这个名字更为大家所熟知。

数字技术也将软件引入到了接收器技术中，这使得使用很便宜的方法构建符合多种标准的设备和可适配系统（如软件无线电）成为可能，这些系统将用在未来的移动通信网络中[Jon02]。

下面我们将仔细分析有关数字接收器的两点内容：带通信号的采样（一个接收器的 ZF 信号如何描述）及信号的正交描述（作为 3.1.6 节的延续）。

1. 带通信号的采样

根据采样定理，连续信号经过正确的采样后采样值能够完整地描述原信号。在这种情况下采样是明确的，即在时间离散的空间内可以可逆地构造出连续性。

采样定理经常引用下面这段话："采样频率必须大于信号最高频率的两倍"。这并没有错，但也不对。为了解释这一原因，可以想象一下带通信号的情况，假设它的频谱位于 80 ~ 100kHz 之间，可以采用两种方法对这个信号进行采样：

1）直接采样：根据上面这段话，采样频率必须大于 200kHz。

2）间接采样：信号首先在模拟空间与 80kHz 的频率进行混频。混频后信号的频率范围是 0 ~ 20kHz。此时采样的速率大于 40kHz 即可。

因为模拟混频过程是不可逆的，这就意味着在混频过程中信息量是不变的。很显然直接采样和间接采样两者的采样序列都表示相同的信息，但两者的区别是数据量相差 5 倍。因此直接采样的采样值必须是冗余的。

上面对采样定理的错误引用描述了很多但并不是必要的明确采样的条件。对低通信号而言上述的条件是必要的，因为低通信号的数字化远比带通信号的数字化频繁，经常会遇到采样定理不能正确描述的情况。式（1-2）的一种较好的表达方法为：

如果采样频率大于模拟信号带宽的两倍，采样是唯一的（即是不可逆的，没有信息丢失）。

采样相当于模拟信号与 Dirac 脉冲序列的乘积。采样信号的频谱变成两个能够折叠在一起的子频谱。Dirac 脉冲序列的频谱（脉冲间隔：T = 采样间隔）还是一个 Dirac 脉冲序列（频率间隔：$1/T = f_A$ = 采样频率），与 Dirac 脉冲序列的卷积意味着周期性的重复。通过采样的模拟信号的频谱也是周期性重复的。现在需要注意的是在这个过程中单个周期性的频谱之间不能相互重叠，多数情况下单个周期频谱之间必须分离。图 5-17描述的是频率范围为 f_u 到 f_o 的带通信号的情况。

在图 5-17 中通过频谱的周期性重复子成员 1 在 f_A 处出现副本，它产生频谱成分 1′、1″和 1‴。成员 2 也会复制出 2′、2″和 2‴。当经过 k 次重复后 1 及 1″的右边缘（陡峭边）与 2 没有重叠时，原始频谱（1 和 2）保持不变。同样 2 的右边缘也不允许与 1‴（1 的第（$k+1$）次重复）重叠。在频谱的负半边这样的关系不可避免的是对称的。用数学的方法可描述如下：

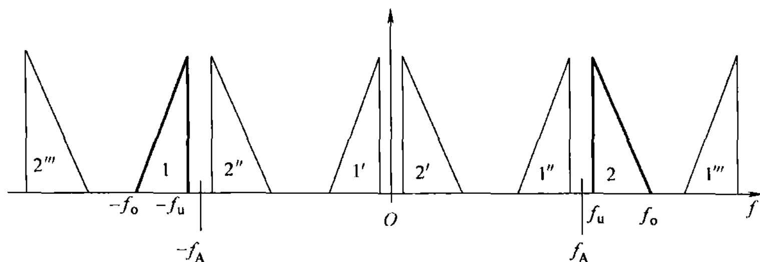


图 5-17　实带通信号的采样（粗线描述的是由成员 1 和 2 组成的原始频谱，细线描述的是经过采样而产生的新的频谱周期）

$$\begin{aligned} -f_u + kf_A &< f_u \\ -f_o + (k+1)f_A &> f_o \end{aligned} \tag{5-9}$$

对f_A进行求解，可得纯带通信号的采样定理为

$$\frac{2f_u}{k} > f_A > \frac{2f_o}{k+1} \tag{5-10}$$

这里 k 是原始频谱内（图 5-17 中块 1 和块 2 之间）周期的个数以及带宽 $B=f_o-f_u$（单边带带宽）在区间 $0\sim f_u$内重复出现的次数。k 必须是一个自然数。根据式（5-10）可得

$$\begin{aligned} (k+1)f_u &> kf_o \\ kf_u + f_u &> kf_o \\ k(f_o - f_u) = kB &< f_u \\ 0 \leqslant k &< \frac{f_u}{B} \end{aligned} \tag{5-11}$$

用数字举个例子结果就会很清楚：带宽为 $B=6\text{kHz}$ 的带通信号的频谱范围是 50～56kHz。根据式（5-11），k 的值可取 0～8。根据式（5-10），采样频率的值是满足条件的，表 5-5 给出了计算的结果。

表 5-5　带通采样的数字实例结果

k	f_A/kHz	k	f_A/kHz
0	112～∞	3	28～33.33
1	56～100	…	…
2	37.33～50	8	12.44～12.50

值得注意的是第 1 行：$k=0$ 意味着已采样信号没有低于f_u的频谱成分。因此它是“正常”的采样，跟处理低通信号时相同。相应的采样频率最小是f_o的两倍，但没有最大值。

此外还值得注意的是 $k=8$ 的另外一种极限情况：最小采样频率大于带宽的两倍。这个解释不仅适用于带通信号，也适用于低通信号，因为后者 $f_o=B$。原因实际上是很

明显的：采样定理涉及的是信息量的守恒。计算信道能力的香农公式中（式（1-17）及式（1-19））出现的变量是 B，而不是 f_o。这可以从上面的公式中推导出来，即将式（5-11）代入到式（5-10）中并求最小值。这个最小值出现在当 k 尽可能大时：

$$f_A > \frac{2f_o}{k+1} = \frac{2f_o}{\frac{f_u}{B}+1} = \frac{2Bf_o}{f_uB} = \frac{2Bf_o}{f_u+(f_o-f_u)} = 2B \tag{5-12}$$

带通采样后一个频谱周期位于低频处，在图 5-17 中就是成分 1′和 2′。这个频谱成分可采用与低通信号一样的处理方法。

对带通信号采样时 A/D 转换器可以工作得较慢。但上游的采样和保持电路（S&H）必须能够按照输入信号的实际频率进行处理。

不太好的是受式（5-10）和式（5-11）的限制不是任何大于 $2B$ 的采样频率都是可用的，因此通过带通采样不是总能够准确地在基带位置进行混频。Hilbert 变换对这个问题会有帮助。

2. 分析信号和 Hilbert 变换

上面所述的带通采样是从实 BP 信号出发的。实时间信号总有共轭的复频谱，周期重复时同样会出现频谱相互重叠的问题。如果时间信号只有单边频谱，那么图 5-17 中的部分频谱就不会出现，式（5-10）的限制也就能放宽：采样频率只要能够满足 $f_A > 2B$ 即可（更精确为：$f_A > B$，但必须对两个信号进行数字化）。

这样的只有单边（“因果”）频谱的信号称之为分析信号。它在时域内是复数值，因为它们的频谱不是共轭复值。分析信号可用两个实函数来表示，其中一个表示复时间信号的实部，另一个表示复时间信号的虚部。这两个函数在分析信号中通过 Hilbert 变换联系到一起。

注意：不是共额复数的频谱都属于复时域信号。分析信号是其中的一个特例，它的频谱不论从哪个角度都是不对称的：如图 5-18 所示，它是单边的。

分析信号的带通采样被称为复带通采样。与此相反图 5-17 所描述的版本为实带通采样。

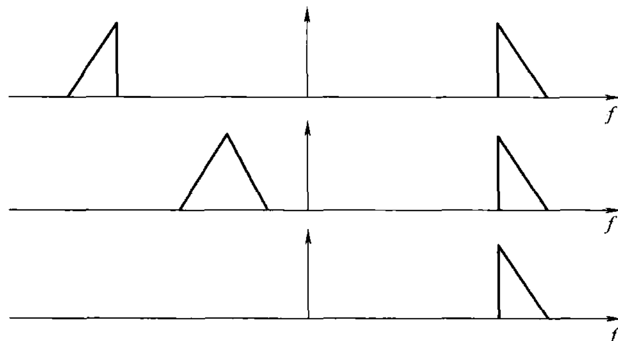


图 5-18　不同的频谱值

上：时时域信号：频谱值是对称的

中：复时域信号：频谱值不对称

下：分析信号：频谱是单边带的

图 5-18 下所示的频谱包含与图 5-18 上所示实信号相同的频谱。实时域信号的频谱是对称的这一点我们已经很熟悉了。现在涉及的是如何从图 5-18 最上面的频谱推导出最下面的频谱。这可以通过两个子频谱的叠加来进行：$X(\mathrm{j}\omega)=X_1(\mathrm{j}\omega)+X_2(\mathrm{j}\omega)$，如图 5-19 所示。由傅里叶变换的叠加定律可得：

$$X(\mathrm{j}\omega)=X_1(\mathrm{j}\omega)+X_2(\mathrm{j}\omega)\multimap x(t)=x_1(t)+x_2(t) \tag{5-13}$$

实信号（例如接收器的 ZF 信号）现在变成 $x_1(t)$。它的频谱 $X_1(\mathrm{j}\omega)$ 是共轭复数，即实部是偶数，虚部是奇数，如图 5-19 最上面所示。现在我们来表示频谱为图 5-19 最下面所示的分析信号。为此我们需要一个频谱为 $X_2(\mathrm{j}\omega)$ 的帮助信号 $x_2(t)$。它的频谱需为图 5-19 中间部分所示的有一个奇数的实部和一个偶数的虚部。傅里叶变换的对称性要求 $x_2(t)$ 必须是一个纯虚部。此外 $x_2(t)$ 必须要能够根据 $x_1(t)$ 计算出来，正如图 5-19 的各部分是比较相近的。因此有

$$x_2(t)=\mathrm{j}\tilde{x}_1(t) \tag{5-14}$$

因此式（5-13）的叠加信号为

$$x(t)=x_1(t)+\mathrm{j}\tilde{x}_1(t) \tag{5-15}$$

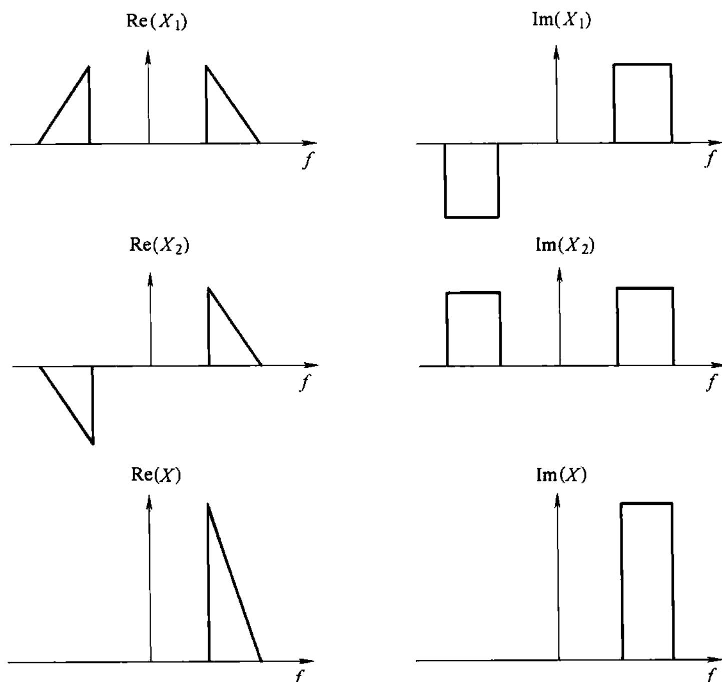


图 5-19　频谱的叠加（Superposition）

现在只需要推导出 $x_1\to X$ 就能将问题解决。这可以很容易地根据图 5-19 在频域内进行表达：

$$X_2(\mathrm{j}\omega)=\mathrm{sgn}(\omega)X_1(\mathrm{j}\omega) \tag{5-16}$$

根据式（5-14）及傅里叶变换的线性性质可得

$$x_2(t)=\mathrm{j}\tilde{x}_1(t) \circ\!\!-\!\!\circ =X_2(\mathrm{j}\omega)=\mathrm{j}\tilde{X}_1(\mathrm{j}\omega) \tag{5-17}$$

根据式（5-17）及式（5-16）即可得到所寻找的变换关系，即 Hilbert 变换：

$$\tilde{X}_1(\mathrm{j}\omega)=\frac{1}{\mathrm{j}}\mathrm{sgn}(\omega)X_1(\mathrm{j}\omega)=-\mathrm{jsgn}(\omega)X_1(\mathrm{j}\omega)$$

$$\tilde{X}(\mathrm{j}\omega)=-\mathrm{jsgn}(\omega)X(\mathrm{j}\omega) \tag{5-18}$$

因此有

$$\begin{aligned}\mathrm{Re}(\tilde{X}(\mathrm{j}\omega))+\mathrm{jIm}(\tilde{X}(\mathrm{j}\omega)) &= -\mathrm{jsgn}(\omega)\mathrm{Re}(X(\mathrm{j}\omega))-\mathrm{jsgn}(\omega)\mathrm{jIm}(X(\mathrm{j}\omega))\\ &=\mathrm{sgn}(\omega)\mathrm{Im}(X(\mathrm{j}\omega))-\mathrm{jsgn}(\omega)\mathrm{Re}(X(\mathrm{j}\omega))\end{aligned}$$

$$\begin{aligned}\mathrm{Re}(\tilde{X}(\mathrm{j}\omega))&=\mathrm{sgn}(\omega)\mathrm{Im}(X(j\omega))\\ \mathrm{Im}(\tilde{X}(\mathrm{j}\omega))&=\mathrm{sgn}(\omega)\mathrm{Re}(X(\mathrm{j}\omega))\end{aligned} \tag{5-19}$$

一个信号 $\tilde{X}$ 的 Hilbert 变换 X 可以通过在频域内交换频谱的实部和虚部得到。

图 5-20 给出一个例子：

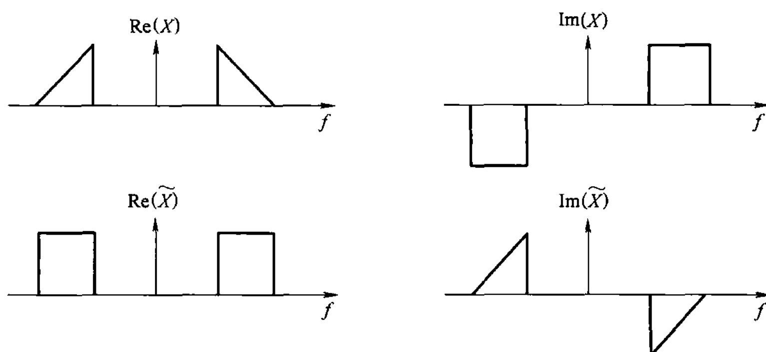


图 5-20 信号的频谱（上）及其 Hilbert 变换的频谱（下）

因为 $\mathrm{Re}(X(\mathrm{j}\omega))$ 是偶函数，$-\mathrm{sgn}(\omega)$ $\mathrm{Re}(X(\mathrm{j}\omega))$ 是奇函数。因此 $\tilde{X}(\mathrm{j}\omega)$ 的虚部是奇数的。对 $\tilde{X}(\mathrm{j}\omega)$ 的实部进行类似的分析可以看出它是偶函数，如图 5-20 所示。$\tilde{X}(\mathrm{j}\omega)$ 的频谱也是一个共轭复数。因此有：

如果 $x(t)$ 是一个实函数，那么 $\tilde{X}(\mathrm{j}\omega)$ 也是一个实函数。

根据式（5-18）可以立即得到理想 Hilbert 变换的频率，通过傅里叶逆变换可以得到冲激响应。

$$H_{\mathrm{H}}=(\mathrm{j}\omega)=-\mathrm{jsgn}(\omega)$$

理想 Hilbert 变换：

$$h_{\mathrm{H}}(t)=\begin{cases}\dfrac{1}{\pi t};t\neq 0\\ 0;t=0\end{cases} \tag{5-20}$$

Hilbert 变换并不改变信号的域，时域信号经过变换后还是时域信号。

图 5-21 是理想 Hilbert 变换的幅度、相位和冲激响应。

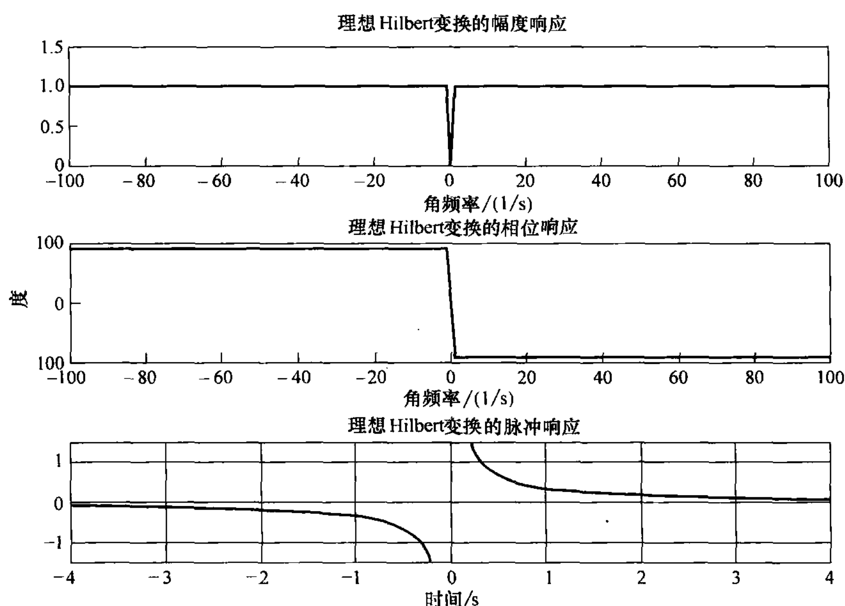


图 5-21　理想 Hilbert 变换的幅度、相位和冲激响应

Hilbert 变换器：
是一个宽带、90°移相器
产生输入信号的正交成分

Hilbert 变换器的输出信号可以通过输入信号与冲激响应的卷积得到。这个积分信号称为

Hilbert 积分：
$$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{x(t-\tau)}{\tau}\mathrm{d}\tau \tag{5-21}$$

对于实现来说，Hilbert 变换器有两大缺点：它有一个无限的带宽和非因果的冲激响应。因此必须采取近似的办法，即使用带宽受限的 Hilbert 变换器。只要被转换的信号的带宽同样是受限的，就不会有任何问题。图 5-22 所示为新的系统函数。

从图 5-22 中不断衰减因此长度有限的冲激响应可以看到，带宽受限的 Hilbert 变换器可以采用 FIR 滤波器很容易地实现。但这种实现只是因果上可行的，时间的偏移（滤波器的包络时延）必须使用图 5-23 所示的电路进行补偿。一个实际的时间信号 $x(t)$ 经过 Hilbert 变换只能实现它的一个经过时延的副本 $x(t-\tau)$。这个时延对绝大多数应用都没有影响。

Hilbert 变换器能看成是一个宽带移相器。这也使上面描述的另一种版本的实现成为可能：使用多个具有频率偏移的窄带移相器（数字全通，即递归系统）来近似上面

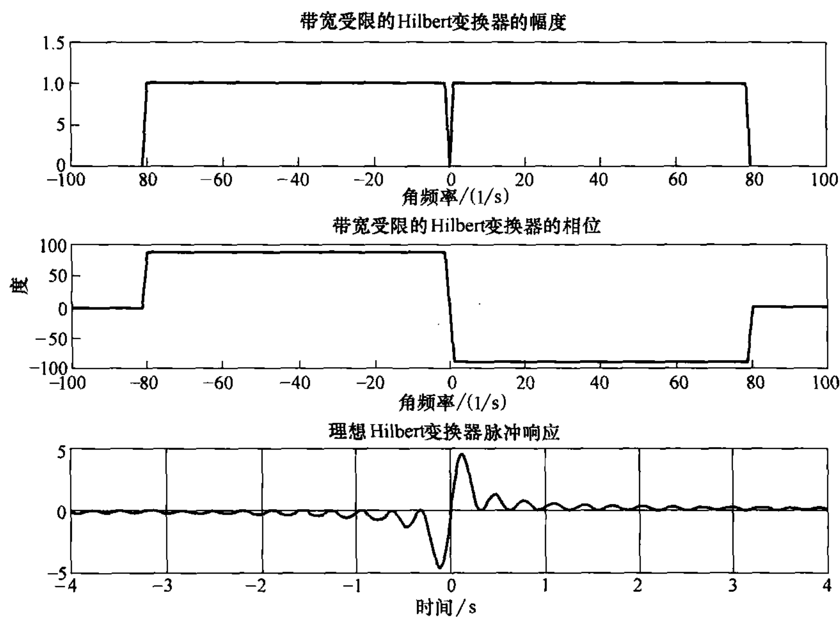


图 5-22　带宽受限的 Hilbert 变换的幅度、相位和冲激响应

的宽带移相器。“FIR 方法”给出一个正确的相位响应和一个近似的幅度响应。“全相位”方法刚好相反。

我们在图 3-18 中就已经遇到了图 5-23 所示的 Hilbert 变换器的实现。那里涉及的是实现 SSB 调制的相位法，Hilbert 变换器用来实现宽带 90°移相器。

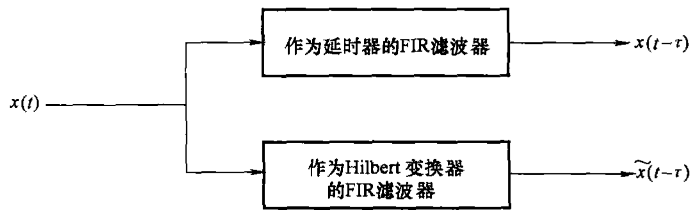


图 5-23　Hilbert 滤波器的实际实现

现在应该说一下 Hilbert 变换的几个有趣的特点。这里使用下面的写法：$\hat{x}(t)=H\{x(t)\}$

1）线性：

$$H\{a_1x_1(t)+a_2x_2(t)\}=a_1H\{x_1(t)\}+a_2H\{x_2(t)\} \tag{5-22}$$

2）时间不变性：

$$\hat{x}(t-\tau)=H\{x(t-\tau)\} \tag{5-23}$$

3）反转性：

$$H\{\hat{x}(t)\}=H\{H\{x(t)\}\}=-x(t) \tag{5-24}$$

相位翻转 90°两次就产生反转。

4）正交性：

$$\int_{-\infty}^{+\infty} x(t)\hat{x}(t)\,\mathrm{d}t = 0 \tag{5-25}$$

5）线性滤波：将 $x(t)$ 和 $\hat{x}(t)$ 通过具有相同冲激响应 $h(t)$ 的滤波器，产生的输出信号 $y(t)$ 和 $\hat{y}(t)$ 仍然是 Hilbert 对应的。

6）对称性：

$$\begin{aligned}&\text{偶信号：}x(t)=x(-t)\rightarrow\hat{x}(t)=-\hat{x}(-t)\\&\text{奇信号：}x(t)=-x(-t)\rightarrow\hat{x}(t)=-\hat{x}(-t)\end{aligned} \tag{5-26}$$

7）相似性：

$$H\{x(at)\}=\hat{x}(at) \tag{5-27}$$

8）能量守恒性：

$$\int_{-\infty}^{+\infty} x^2(t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} \hat{x}^2(t)\,\mathrm{d}t \tag{5-28}$$

9）调制特性：

$$H\{s(t)\cos\omega_0 t\} = s(t)\sin\omega_0 t \tag{5-29}$$

前提条件：$s(t)$ 在频率 $|\omega_0|$ 处的带宽是有限的。

10）几种对应关系：

表 5-6　Hilbert 变换的几种对应关系

$x(t)$	$\hat{x}(t)$	共振条件
$\cos(\omega_0 t)$	$\sin(\omega_0 t)$	$\omega_0>0$
$\sin(\omega_0 t)$	$-\cos(\omega_0 t)$	$\omega_0>0$
$\delta(t)$	$\dfrac{1}{\pi t}$	无
$\dfrac{\sin(\omega_g t)}{\omega_g t}$	$\dfrac{1-\cos\omega_g t}{\omega_g t}$	无

将式（5-29）与式（3-31）和式（3-32）的导出式相比可以看出，一个信号的同相和积分成分不外乎就是 Hilbert 对应物。式（3-28）描述的复包络就是一个分析信号。

复信号也被用来进行理论上的观察。一个实 BP 信号可以用等价的分析信号来描述，它们又可用等价的 TP 信号来描述，如图 5-24 所示。因此足够可以找到将 TP 信号和系统用于 BP 信号和信道传输的理论。

总结：采用单边频谱来对（特别是数字处理）已调信号进行处理是有好处的。它们在时域上是复数，虚部是实部的 Hilbert 变换。分析信号的处理要求有两条信号路径，每条都是对一个实信号进行处理。

傅里叶变换的频率偏移（也被称为调制）之后紧跟着就是正交混频：

$$x(t)\mathrm{e}^{\mathrm{j}\omega_0 t} \circ\!\!-\!\!\bullet\, X(\omega-\omega_0) \tag{5-30}$$

式（5-30）的时间信号可以是复数值的。选择分析信号是有益的，这样在混频前

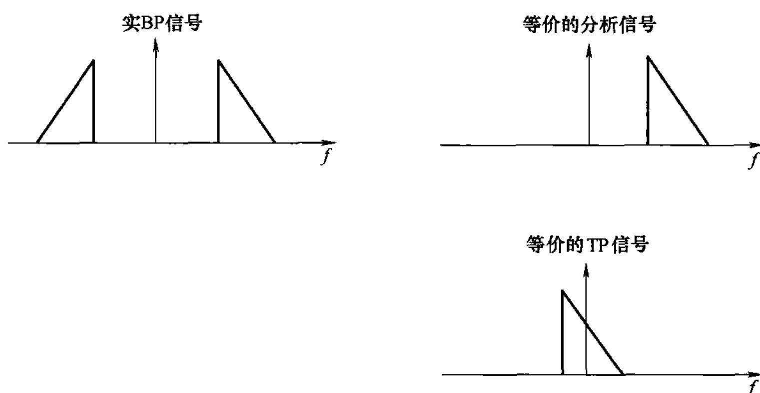


图 5-24 等价 TP 信号的推导

后它都是唯一的，即它不会产生镜像频率。

$$\underline{x}(t)=x(t)+\mathrm{j}\hat{x}(t) \tag{5-31}$$

因此正交混频过程为

$$\underline{y}(t)=y(t)+\mathrm{j}\hat{y}(t)=\underline{x}(t)\mathrm{e}^{\mathrm{j}\omega_0 t}=\underline{x}(t)[\cos\omega_0 t+\mathrm{j}\sin\omega_0 t] \tag{5-32}$$

由此可以得出图 5-25 所示的原理图。复值信号采用两个平行的箭头来分别描述实部和虚部。

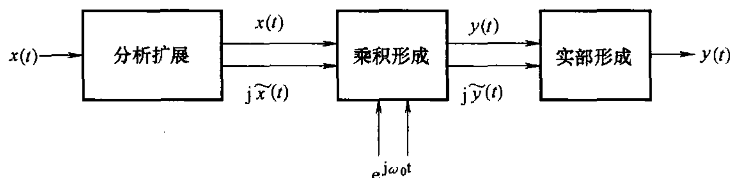


图 5-25 正交混频的原理

实输出信号跟式（3-36）相同通过构造 $y(t)$ 的实部来得到。由式（5-32）可以得出

$$\begin{aligned}y(t)&=\mathrm{Re}(\underline{y}(t))=\mathrm{Re}([x(t)+\mathrm{j}\hat{x}(t)][\cos\omega_0 t+\mathrm{j}\sin\omega_0 t])\\&=x(t)\cos\omega_0 t-\hat{x}(t)\sin\omega_0 t\end{aligned} \tag{5-33}$$

图 5-26 给出了相应的结构图，它最终可使用实信号路径来处理，并且由于使用延时器（比较图 5-23）它是因果可实现的。

对于一个频谱成分位于较高的中间频率 ω_0 的窄带信号（例如一个已调载波信号），调制公式（5-29）是成立的。对这些信号，可以据此将乘法器之前的分析扩展省略掉，于是有图 5-27 所示的简化结构图。与图 5-26 不同，同相和正交成分并没有在这里加，对它们要进行后续的单独处理。混频器是传统的构造类型，它能产生差频与和频，但后者被低通滤波器给过滤掉了。低通滤波器后的两个信号 $y(t)$）和 $\hat{y}(t)$ 表示（可能是已调）信号 $x(t)$ 的复包络。这就是 $x(t)$ 的等价 TP 信号（见图 5-24），它包含已调

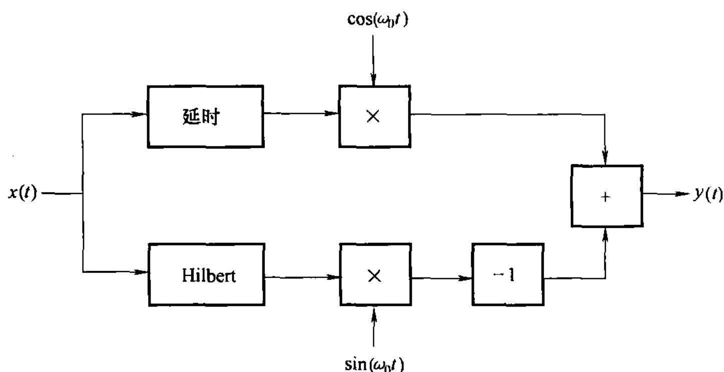


图 5-26　正交混频器的实际实现

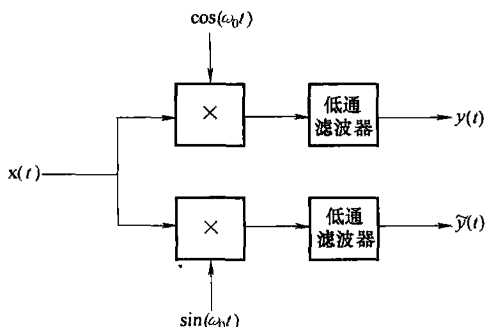


图 5-27　形成复包络的窄带信号的向下混频器

信号 $x(t)$ 的所有信息，这在 3.1.6 节中已经进行了推导。解调可以使用复包络单独进行，使用纯数字的信号处理器就能完成。例如式（3-3）就描述了一个 AM 信号如何进行不相干解调（包络检测器）。

现在我们可以思考数字接收器的构成原理了。图 5-28 所示为使用重叠原理的接收器。HF 和 ZF 部分与图 5-15 的结构相同。有时会有第三个 50kHz 以下的中间频率，目的是能够插入一个便宜的 A/D 转换器。由于频率较高，HF 和 ZF 部分仍然使用模拟的方式实现。相反电话调制解调器直接对到来的信号进行数字化，只包括 A/D 转换器和图 5-28 中没有框起来的部分。

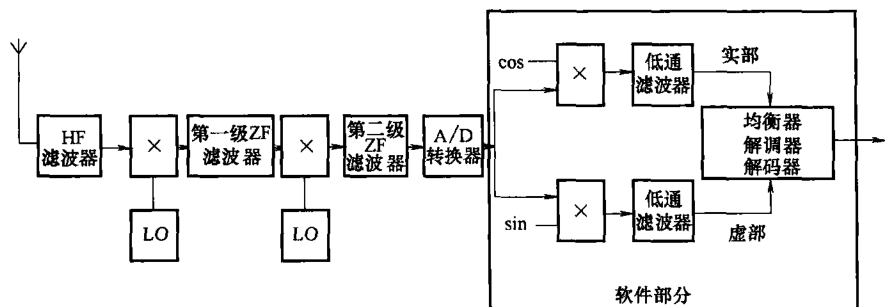


图 5-28　依据重叠原理、带有数字解调器的接收器

（带通位置的实 ZF 信号）（没有画出放大器）

图 5-29 所示为使用混频原理的接收器。这里根据图 5-27 构成了一个复杂的 ZF 信号并直接在基带进行混频。

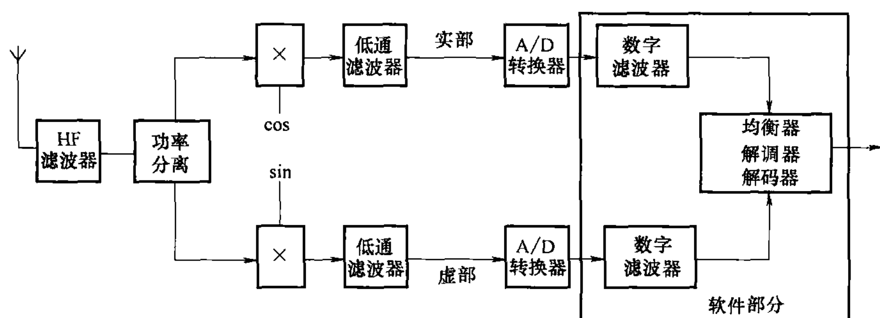


图 5-29　依据直接混频原理的接收器（基带的复 ZF 信号）和数字解调器

发送器采用同样的方式实现，即调制使用 DSP 在等价基带上实现，然后使用复数的方式搬移到带通上。在那里还要构造一个实数部分，比较图 5-25。

利用分析信号和正交描述的背景知识还有两个有趣的细节需要解释：

1）在 3.4.4 节中描述了数字相位调制（PSK）。解调器需要相位正确地再生载波（相干解调）。图 3-8 所示的二次方运算是载波再生的一种可能方法。

那里经常使用的一种方法是 Costas 环。它无外乎就是一个具有反馈控制功能的正交解调器。反馈能消除正交成分，因此同相成分与载波的相位是相等的。

2）一个分析信号 $\underline{x}(t)$ 具有单边的（因果）频谱。条件是 $\underline{x}(t)$ 的虚部是实部的 Hilbert 变换：

$$\underline{x}(t) = x(t) + \mathrm{j}\hat{x}(t)$$

由于傅里叶变换与反变换是对称的，相同的重叠在频域内也是可以做到的：

> 如果一个时域信号是因果的（所有的冲激响应可实现的系统都满足这个条件！），那么该信号的频谱（系统的频率响应）的虚部必须是实部的 Hibert 变换。

因此所谓的最小相位系统（它的传输函数在右半 s 平面及 z 平面的单位圆之外没有零点）会遵循：

> 最小相位的线性系统的幅度响应和相位响应不能独立地相互选择！

5.3　Spread-Spectrum 技术（扩频技术）

根据图 4-31，增大带宽是提高信道能力的一种可能方法。这可以通过信道编码、调制或 Spread-Spectrum 技术（扩频技术）来实现。Spread-Spectrum 技术和调制的主要区别是扩频技术并不是对信息本身，而是对一个特殊的扩频序列的带宽进行增加。

这种技术在 20 世纪 30 年代中期就为人所知，但在 70 年代末才在除军事领域之外的领域得到实际应用。目前 Spread-Spectrum 技术也用在民用上，主要是卫星的接入方法，参见 6.2.1 节。

大致有 3 种不同的 Spread-Spectrum 技术：Direct sequence spread spectrum（DSSS）、Frequency hopping spread spectrum（FHSS）、Chirp-spread-spectrum。

所有这 3 种方法都能在不改变信号发送功率的前提下增大 HF 频谱。如图 5-30 所示，功率谱密度却明显地减小了。例如图 5-30 中的信号 A 可能是传统的 FM 信号，信号 D 被扩展得很厉害，以至于它的功率密度很低。使用正常的接收器根本检测不到这个信号。加密只能隐藏信息的内容，使用 Spread-Spectrum 信息甚至可以自己传输。因此 Spread-Spectrum 主要对军事应用很有吸引力。

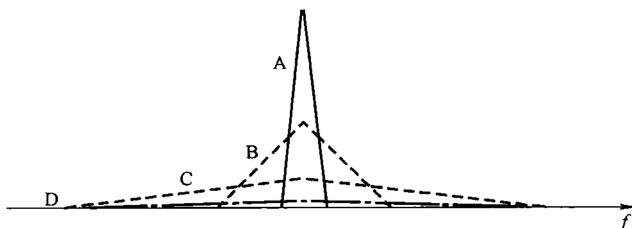


图 5-30　不同扩展系数下发送功率的频谱分布

在 frequency-hopper 的情况下，载频使传统发送器多样化。使用随机产生器可以从例如 100 个频率中选择一个并在几毫秒之内占用。接收器必须知道这个随机序列，并且本地振荡器要紧跟发送器跳变。在接收器的 ZF 部分中的信号与传统发送器的信号相同。

与传统的传输相比它有如下的优点：

1）如果一些频率已经被传统信号占用，接收器还可以接收例如 90% 的信息。由于使用信道编码或存在话音的冗余，信息的内容不会受到损伤。相反在传统的传输中如果利用已经被占用的信道，信息会完全被毁掉（例如在 FM 系统中存在攫取效应）。

2）非希望的接收器不知道这个随机序列。使用扫描接收（自动寻找接收器）只能接收到部分的信息，这能够保护信息的内容。

3）从技术上来说干扰发生器不能使用一个干扰信号来覆盖例如 50MHz 的带宽。多数情况下干扰器用脉冲驱动攻击陌生的、在短暂发送间隔内通过快速扫描自己搜索到的发射频率。frequency-hopper 可减弱这些干扰。这也是这种技术在军事上得到应用的主要原因。

在 HF 频谱平均较短的持续时间内，根据是否检测到发射频率，frequency-hopper 可能有较高的功率密度或功率密度为零。只有通过较长的时间才会出现较小的功率密度。因此传输本身不是隐蔽的。同样 frequency-hopper 也很容易被发现。对它的定位（不用对内容进行评估，总结出几个百分比的概率即可）甚至比对传统的发送器定位还容易，因为信号是基于不同的频率的，所以只要计算出地形的中间值即可。

地形会改变吸收和反射关系，因此希望的接收器会发现不同的瞬时频率下接收关系是不同的。接收器的 AGC 必须能够很快地做出反应。因此在山区频率跳变的速率限制在每秒 200～300 次。跳频器的主要应用在于防御电子干扰，只有在需要时才激活跳频运行状态。

图 5-31 所示的 DSSS 对跳频方法有所改进。它产生一个宽带频谱，在这个频谱上

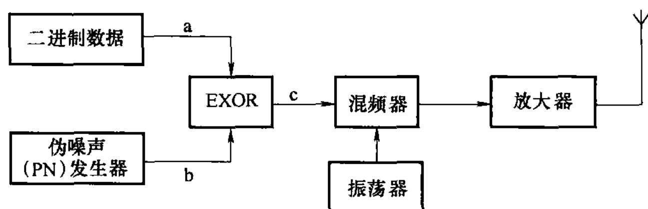


图 5-31　DSSS 发射器的结构图

短时测量功率密度也很小。信号甚至可以存在于接收器的噪声中。

回顾一下图 3-63 可以看出，二进制序列在图 5-31 中的 c 点是 PSK 调制的。它产生一个 $\sin(x)/x$ 频谱，它的带宽与序列 c 的时钟速率有关。这个时钟速率会人为地大幅度提高，这样 PN 发生器的时钟速率才能是信息源数据速率的多倍。PN 发生器产生一个伪随机序列 b（PN），也就是说序列是周期性出现的，但这个周期很长，在一个周期内它具有随机性。这可以很容易地用带反馈的移位寄存器来实现。扩频序列 b 的符号也被称为 chip。

接收器用一个混频器来对序列 c 进行解调。此外它还用另一个 PN 发生器产生相同的扩频序列 b 并通过一个异或门 EXOR 与 c 相连。异或门的输出就是原始序列 a。两个相连的异或门 EXOR 能恢复出原始信号，这我们在 4.2.2 节中的连续加密中已经看到了。那里两个序列具有相同的时钟速率，这里扩频序列明显快很多。接收器的主要问题是它的 PN 发生器需要与发射器中的 PN 发生器同步。

不希望的传统接收器很少能检测到扩频信号，不希望的扩频接收器不知道扩频序列，不能对信号进行压缩。

窄带干扰信号经过扩频接收器后被扩频，如图 5-32 所示，因此不会产生影响。

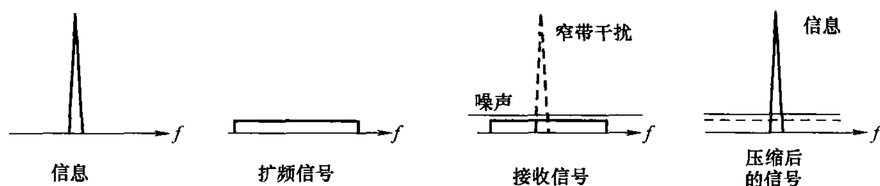


图 5-32　直接序列扩频传输的信号频谱

不希望的接收器将扩频信号（如果它是扩频信号的话）当成是小噪声的增加。可以将已扩频的信号同时在相同的频率范围内传输，相互的干扰会使信噪比下降。这会引出作为 FDM 和 TDM 变种的 Code Division Multiplex（CDM）。

FDM 情况下，多个发送器同时工作在不同的频率上，TDM 时多个发送器在不同的时间工作在同一频率上，CDM 情况下多个发送器同时工作在相同的频率上，如图 1-36 所示。

CDM 的前提条件是在单个传输之间有较少的干扰，正交扩频序列可以满足这个要求，扩频序列的作用相当于寻址。不同用户之间的协调只是一次性的，无需在运行之

间进行。正是因为这个原因 CDM 对卫星系统是很有吸引力的。UMTS（通用移动通信系统）也使用 DSSS，蓝牙（房间内的无线通信系统，可以不用使用 PC 电缆）使用 FHSS。

5.4　PLL 与频率综合

5.4.1　锁相环

PLL 实际是一个正常的反馈环。但下面两方面原因使得它的实现还需要一些努力：

1）被调节的大小是个相位（直觉就比较困难）；

2）在一定的运行环境下 PLL 具有很强的非线性（数学描述很困难）。

在这里才讨论 PLL 的原因是它在信息技术中多方面的应用。

PLL 由 3 个单元组成，即：VCO（Voltage Controlled Oscillator，压控振荡器）、PD（相位检测器）、LF（环路滤波器）。

这些组成单元按图 5-33 的方式连接起来。

PD 比较输入信号 u_1 和 VCO 输出信号 u_2 的相位，并给出是相位差的百分比的信号 u_d。这个信号通常含有 AC 成分，它将被 LF（低通滤波器，通常是一阶的）去除。信号 u_f 的变化很慢（理想情况下它只包含 u_d 的 DC 部分），它来控制 VCO。

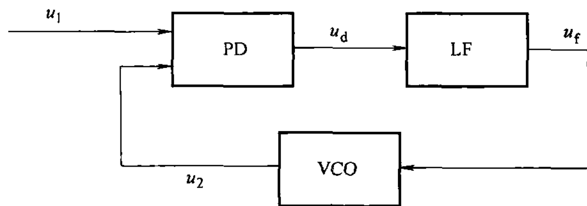


图 5-33　PLL 的组成结构

如果 u_2 的相位落后于 u_1 的相位，u_f 就会增加，u_2 的频率也会随之增加。如果 u_2 的相位超前于 u_1 的相位，u_f 就会下降，u_2 的频率也会随之降低。在锁定情况下（只有此时 PLL 才是可用的）u_1 和 u_2 的相位差为常数，并具有相同的频率。

PLL 的应用是多方面的，前面已经提到过几个，下面列出一些其他可能的应用：数据接收器的时钟再生器、解调器的时钟再生器、减小抖动器、移相器、频率同步、相位同步、FM 和 FSK 解调器、PM 和 PSK 解调器、倍频器和分频器（5.4.2 节）、频率合成器（5.4.2 节）及其他。

因为有这么多的应用，所以已经存在许多不同的 PLL 实现方法是不足为怪的。可以将这些方法归纳成 4 类[Bes93]。

1）线性 PLL：图 5-33 所示的带有模拟 LF、复用器或环形混频器作为 PD 以及具有连续可设置输出频率的 VCO 的 PLL 就是线性 PLL。

2）古典数字 PLL：PD 是用数字的方法实现，如采用异或门。其他构成单元同线性 PLL 相同。该 PLL 具有模拟-数字混合电路。

3）全数字 PLL：VCO 使用 DCO（Digitally Controlled Oscillator）来实现，LF 也是纯

数字的实现。

4）软件 PLL：所有的操作都是通过软件在一个 DSP 上来实现。目前这种方法只能用在慢速应用中。

5.4.2 频率综合

在信息技术系统中经常会用到振荡器，来为混频器、调制器和解调器提供谐波。根据使用的领域不同，对这些振荡器的评价指标如下：频谱的纯洁度，温度变化时频率的稳定性和负载的变化性及老化，可选择的频率范围，频率设置的精确度（解析度），频率调节或变化的速度，大小，价格。

显然，没有电路能够同时满足上面的所有的要求。因此在实际中人们会使用不同概念的振荡器。振荡器的开发也是一直在不断地寻求各方面最优值的折中。通常也会将不同的振荡器结合起来使用，以便统一不同方面的优点。

最早使用的是 LC 振荡器，它由串行或并行的共振环路构成。这些振荡器可以提供纯谐波信号，频率简单地通过改变电导或电容来设置。使用电子电路这些是可以做到的，例如可以利用一个变容二极管（varicap）。它的缺点是频率稳定性不足，并且它使用了线圈。频率很高使这些振荡器不易控制。

如果频率不是可变的，可以使用晶体振荡器（Crystal Controlled Oscillator）。它是一种机械的振动器，工作在 MHz 频段上，它的频率稳定性通过调节温度还可以提高。晶体振荡器的缺点是它的频率是不可调的，并且频率范围是受限的。

例如在 PLL 中使用的 VCO，就是一个晶体振荡器，它的频率很容易使用电子电路的方法来设置，但频率的稳定性不是很高。

很高的频率（微波范围）可以按照 1.2.3 节讲述的通过倍频的方法来产生。如图 5-34 所示，这个过程充分利用了非线性。输出频率具有相同的相对精确性，稳定性同晶体振荡器一样。经过倍频的绝对非精确性根据应用的不同可能会导致一些困难。

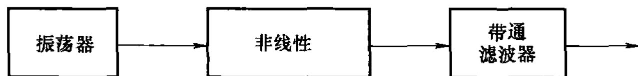


图 5-34 非线性倍频

VFO（Variable Frequency Oscillator）的频率范围是受限的。使用如图 5-35 所示的电路，通过利用多个较小的子范围，会使 VFO 的频率范围增大。

目前大多数的振荡器利用一个甚至多个 PLL 电路，如图 5-36 所示。这里的参考振荡器可能是温度稳定的晶体振荡器，如果有极特殊的要求它还可以是一个铯原子钟。

PLL 调节 VCO 以便使相位检测器输入端的两个信号的相位固定，因而保持相同的频率。由于有分频器，VCO 的频率比参考频率提高了 N 倍。

使用 PLL 也可以很容易地利用一个不稳定的 VCO 来实现一个稳定的、具有数字可选频率的振荡器。

如果希望图 5-36 所示电路的输出频率可以在更密的信道栅格（例如手持移动设备是 10kHz）内可选，参考频率必须较低。但晶体振荡器却不能工作在这么低的频率上。

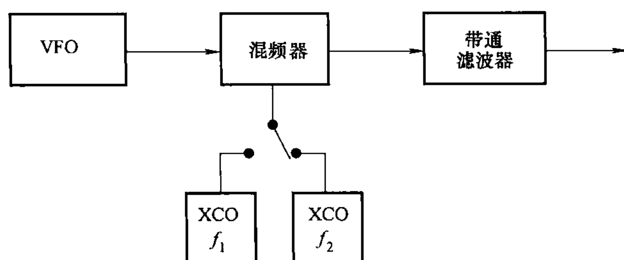


图 5-35　有较大频率变换范围信号的产生

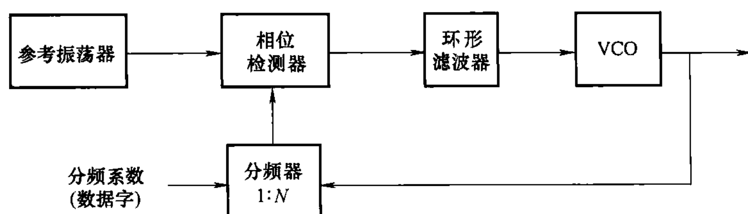


图 5-36　用 PLL 进行倍频

解决办法是在图 5-36 的参考振荡器和相位检测器中间放一个分频器，如图 5-37 所示。

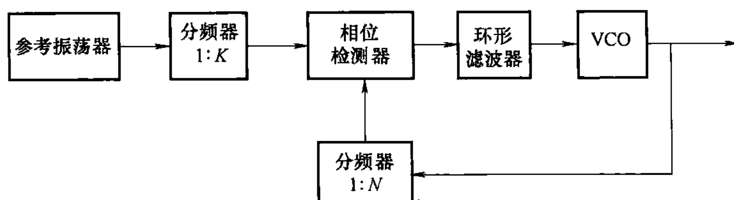


图 5-37　输出端具有较细频率栅格的 PLL

当然人们希望尽可能多的功能能用数字的方式来实现，并且各组件能够高度集成，但使用省电的 CMOS 技术不能实现超过 100MHz 的频率，图 5-38 所示的方法对缓解这个问题有所帮助。图 5-36 或图 5-37 的分频器被分成一个分频器和一个预分器（prescaler）。VCO 和预分器可以具有很高的频率（微波，可以通过 ECL 或砷化镓 CMOS 电路获得（frequency synthesizer IC））。唯一特殊的单元是参考振荡器需要一个振荡晶体。使

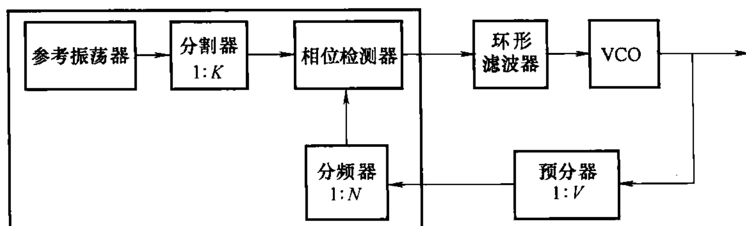


图 5-38　高频 PLL 振荡器

用预分器来代替分频器能利用混频的办法降低 VCO 的频率。VCO 的输出信号也可以通过倍频的方法继续将频率升高。

图 5-38 所示的概念也有缺点：频率解析率值不是约为$\frac{f_{\mathrm{Ref}}}{K}$，而是$\frac{f_{\mathrm{Ref}}}{K}V$，且 V 大于 1，否则预分器是无意义的。

作为补救措施当然是将几种方法结合起来，使用一个图 5-37 所示的具有细解析度的 PLL 来代替图 5-35 中的 VFO。不使用多个晶体振荡器，而用另一个粗粒度的 PLL 来代替即可。带通滤波器后频率由下面公式给出：

$$f_{\mathrm{out}} = N_1 f_{\mathrm{Ref}_1} + N_2 f_{\mathrm{Ref}_2}$$

两个 PLL 的参考频率都由一个共同的模板振荡器得到（出于开销的原因它无论如何都是值得借鉴的），可以选择例如 $f_{\mathrm{Ref}_1} = 10 f_{\mathrm{Ref}_2}$，这样可以得到

$$f_{\mathrm{out}} = f_{\mathrm{Ref}}(10N_1 + N_2)$$

图 5-39 所示为它的原理。利用这种原理可以用多位数字得到希望的频率。如果有人愿意的话可以以图 5-34 为例设计一个倍频器。

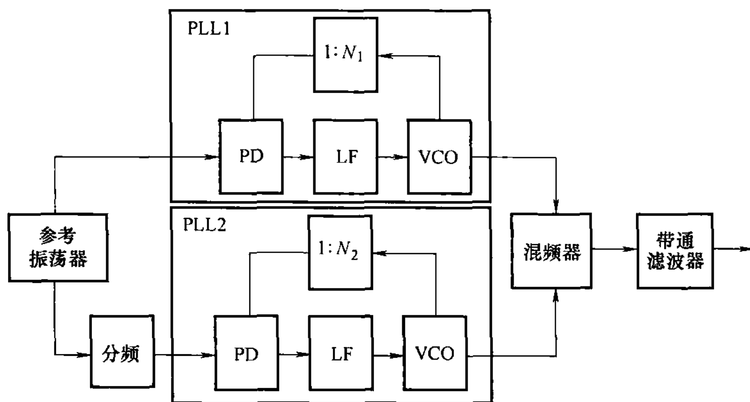


图 5-39 图 5-35 之离散频率版本

PLL1 使用粗粒度频率栅格，PLL2 使用细粒度频率栅格

但图 5-39 的复杂性有点高，因为多个 PLL 相互作用，并且运行着多个相位检测器、环路滤波器及 VCO。最主要的是带通滤波器在混频器之后还必须是可调节的。

使用一点技巧就可以用一个 PLL 来实现同样类型的频率合成器，但这里不对这个技巧继续讨论（具体方法可参见参考文献【Bes93】）。在图 5-38 中用一个可设置的预分器来取代简单的预分器。它可以选择分频系数为 V、$V+1$、$V+2$ 或 $V+3$。这样的电路可以通过集成“带有预设置的模 C 计算器”来获得。此外不是使用单个的分频器，而是使用多个轮流工作的分频器。使用这种方法可以构建一个具有任意大频率范围、任意小解析率、任意好稳定性的合成器。整个电路是可以集成的，看起来像图 5-38（带有复杂的预分器和分频器），受 IC 技术的影响最高工作频率受限。但这个范围通过使用砷化镓技术的独立预分器或是混频器或是倍频器可以继续扩展。

通常情况下需要一个谐波输出信号，但上面介绍的使用 VCO 的数字方法产生的通常是脉冲形式的信号。使用能提取出方波的基波成分的低通滤波器可改变上面的方法。但这种方法并不值得推荐，因为滤波器的极限频率必须不断根据输出频率进行调节。只要频率足够低，图 5-40 所示的方法就不失是一种好方法。在 EPROM 中存储着正弦信号的值，模 k 的计数器作为地址生成器，可对存储器的所有单元进行访问。如果正弦波的一个周期用 2^k 个采样值存储着，那么计数器的时钟必须是输出信号的 2^k倍。这个时钟频率可用前面介绍的频率综合器来产生。使用频率合成器来产生相应的时钟频率，这个电路也可以实现 FM 或 FSK。

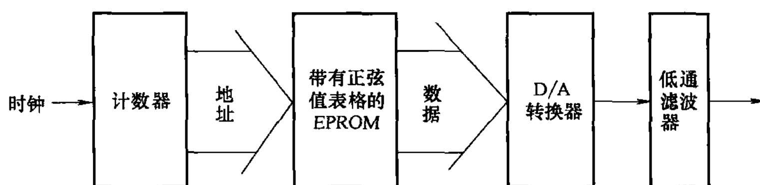


图 5-40　纯谐波信号的数字实现方法

如果只存储一个象限的正弦波的值，那么图 5-40 所示的方法就会节省许多存储空间，其他象限的值可以通过对称的方法得到。此外使用相同的计数器可以同时读出存有正弦值和余弦值的两张表格。这样可以产生相位绝对固定、与频率无关的正交混频器（见图 5-26、图 5-27 及图 5-29）的输入信号。QAM 调制器和 QAM 解调器也使用这种电路，如图 3-21 和图 3-72 所示。

图 5-40 所示方法的关键在于输出端谐波信号的采样频率。当频率变化很大时必须跟踪在输出端口的重构的低通滤波器。已经得到广泛应用的 DDS（Direct Digital Synthesis）方法对解决这个问题很有帮助。

DDS 的基本思想是所有频率都可以看成是相位随时间的变化，即

$$\omega = 2\pi f = \frac{d\varphi(t)}{dt} \approx \frac{\Delta\varphi(t)}{\Delta t} = \frac{\Delta\varphi(t)}{T}$$

在数字系统中最小的时间间隔就是采样时间 T。用图 5-41 所示的电路就可以产生上述形式的频率。

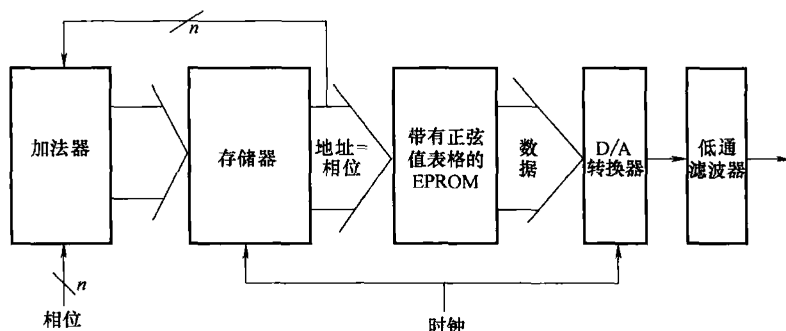


图 5-41　根据 DSS 原理进行频率综合

在 EPROM 中同样也使用了正弦值表格。存储器的地址作为正弦函数的参数（角度），EPROM 的输出端就是与之对应的瞬时值。这些值供 D/A 转换器使用，并被重构的低通滤波器平滑。这部分电路就相当于图 5-40 所述的概念，它们的重要区别是这里时钟的频率是稳定的。因此重构低通滤波器保持不变。

频率会随着相位的增加而改变。对低频来说，相位的增加较小，输出端的正弦函数由很多的相关值来合成。频率较高时相位的增加很大，但正弦值表格只包含少数的相关值。只要采样定理还能够被满足（到时钟频率的一半），就能重新获得正弦函数。频率的稳定性由时钟发生器的稳定性来确定，但正如前面提到的，后者始终在同一个频率上摇摆，因此相对容易做得比较稳定。频率的解析度由存储器地址的位宽来确定（实际中地址总线需要 24 位或 32 位）。输出信号的纯度由 D/A 转换器的字宽来决定（实际中地址总线需要 8 位或 12 位）。最大输出频率由所使用的各逻辑单元来确定。

用度来表示的相位的增加为： $\Delta\varphi = \frac{f_{\text{out}}}{f_{\text{Takt}}} \times 360°$

频率的解析度为： $\Delta f_{\text{out}} = \frac{f_{\text{Takt}}}{2^n}$

数值实例：当时钟频率为 33.5MHz，总线位宽为 $n = 24$ 时，频率解析度为 2Hz，频率范围为 15MHz（理论值为 16.75MHz）。

对 DDS 发生器来说，也有特殊的集成电路。当传输频率较低时当然也可用 DSP（数字信号处理）采用软件的方法来解决。目前输出频率限制在约 40MHz。对需要更高频率的应用来说还得再使用倍频器或混频器。

正交信号可以使用两个表格来产生。由于有数字电路及可以很快地改变频率的设置，DDS 发生器对所有类型的调制器也足够适用。FM 调制器根据信息的内容来改变相位的增加，相位调制器给图 5-41 所示的 EPROM 的输出端增加一个可变的相位。AM 可以通过让图 5-41 所示的 D/A 转换器的参考电压由另外一个受信息内容控制的 D/A 转换器来提供而生成。在 D/A 转换器之前数据总线的数字复用器也有同样的效果。因为 AM、FM 和 PM 是可同时实现的，所以 3.1.6 节所希望的所有的有关调制的内容都可以实现。

图 5-39 所示的 PLL 综合器同样能提供较高的频率解析度。但这些频率不能进行有效的改变，因为考虑到输出信号频率纯度的原因，环路滤波器的带宽必须很窄。显然，在这里 DDS 原理具有优势。

对很多应用来说振荡器信号的频率纯度是很重要的。所有实际振荡器都有幅度和相位波动问题（后者也能引起频率波动）。正弦信号的每次改变实际都会引起无数的很窄的频谱，频谱分析表明多余的频率根据来源不同会变成幅度噪声（amplitude noise）或相位噪声（phase noise）。多数情况下，幅度噪声跟相位噪声相比可以忽略。

因为相位噪声是一连串的随机信号波动，它会产生由幅度密度确定的连续频谱。因此振荡器输出信号要使用高级的频谱分析仪进行测量，结果在频率轴上用 dBc/Hz 来表示。参考值 0dBc 是载波（carrier）功率。因此载频（偏移是 0Hz）的幅度密度就是 0dBc。测量曲线离开载频越快，振荡器信号的频谱越纯。它是用来区别调制带宽内的

带内噪声（in-band noise）和带外噪声（out-of-band noise）的。

特别对无线传输的调制和解调来说，振荡器信号的频谱纯度是很重要的。在发送器端，带内噪声会引起信噪比的下降，因此会导致传输质量降低（固有噪声）。相反，带外噪声通过不必要地加宽自己的发送信号会引起相邻信道的干扰（陌生干扰）。

在接收器端两种类型的相位噪声都会引起信噪比的下降，因而会影响接收器的动态范围。这是因为邻居信道较强的接收信号在混频器中与较弱的振荡器信号进行调制，因此会产生一个很宽的频谱。这会增加接收器的噪声从而损害希望接收到的信号。

前面已经提到过在每种类型的倍频过程（应该是使用非线性和使用 PLL）中绝对频率误差也会翻倍。对频率解析度较高的振荡器来说 DDS 方法与图 5-37 所示的 PLL 方法对相位噪声考虑得同样多。

现代接收器对长波到短波波段范围为 100kHz ~ 30MHz 的频率可以提供 1 ~ 10Hz 的解析率。考虑到较好的镜像频率抑制功能，第一级中间频率选择较高的，例如 45MHz（见表 5-4）。这样第一级本地振荡器必须产生 45 ~ 75MHz 的频率，这对一个 DDS 来说（还）太高。因此通常将一个有较差解析度、有较小相位噪声的 PLL 与一个具有较高解析度同样具有较小相位噪声的 DDS 振荡器结合起来使用。结果就是图 5-39，只不过这里 PLL1 被 DDS 来代替。

5.5 无线传输

Maxwell 在 1865 年预言了电磁波，之后 Herz 在 1888 年验证了它确实存在，Marconi 第一次从技术上使用了它。1901 年 12 月 12 日，Marconi 成功地利用电磁波沟通了大西洋两岸。

Maxwell 之后有了电场和磁场，它们分别用矢量 $\vec{E}$ [V/m] 和 $\vec{H}$ [A/m] 来表示。Maxwell 公式指明，随时间变化的电场能产生磁场，相反，随时间变化的磁场也能产生电场。这两种场可以相互产生，一起构成了电磁波。电磁波能传输能量，并且根据式（1-24）也能传输信息。在自由空间中，即在无干扰的传播情况下，电场和磁场相互正交，瞬时功率流的值和方向根据下面的矢量积来计算：

$$\vec{S} = \vec{E} \times \vec{H} [W/m^2] \tag{5-34}$$

$\vec{S}$ 是 Poynting 矢量，单位是功率密度。根据式（5-34），电磁波与水波一样是个横波，即场的移动（水滴）与波的传播方向是垂直的。

$\vec{E}$ 场与 $\vec{H}$ 场的幅度之比就是波阻抗。在真空中和在空气中它都是实数，值为

$$Z_{W_0} = \frac{|\vec{E}|}{|\vec{H}|} = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi\Omega = 377\Omega \tag{5-35}$$

在真空和在空气中的传播速度为

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = c = 3 \times 10^8 \text{m/s} \tag{5-36}$$

这个结果后来引出了已被证明的假设，即光也是电磁波。

电磁波的极化在 **E** 矢量的方向上。人们会区分：

（1）线性极化

1）垂直极化；

2）水平极化。

（2）圆形极化和椭圆形极化

1）左旋转；

2）右旋转。

电磁波在真空中也能传播。这在 20 世纪还很难想象。因此人们发现了以太，即一种很轻的观察不到的媒介，它使光在真空中的传输称为可能。尽管目前在口语中还有这种说法，但 Albert Einstein 抛弃了对以太的想象。“funken（广播）”这种表达方法也是一样：第一个发送器是 funken 发送器，因为它能从被删除的 Funk 的宽带频谱中产生出高频。当第一次用振荡器实现这个原理时，人们非常高兴。但“funken”这种表达最终也存活了下来。

Maxwell 的 4 个公式以简洁的方式描述了电子力学。但它涉及的是差分公式，只对简单几何数组可解。对物理学家来说 Maxwell 公式是理想的，因为它有巨大的说服力。相反对于工程技术人员来说，它的这种极端的表达方法更像一个应用，而不像对自然的理解。因为 Maxwell 公式比较难解，工程技术人员不太喜欢它，他们更喜欢简单且在数学上比较容易处理的理论。但简化的理论只能针对特殊情况。不过只要它们出现的足够频繁，就值得提出简化的理论：

1）低频范围：这里电磁波的波长大于所观察的系统的尺寸，因此不用考虑波的传播效果。因此人们使用振荡而不是波来工作。同样在低频范围内没有必要使用矢量值如 $\vec{E}$ 和 $\vec{H}$ 来工作。使用由此而导出的标量值如电压 U、电流 I 和通量 ϕ 等的操作是很方便的。

2）100MHz 以下的频率：波长的值最小为 2m（电缆）和 3m（空气）。UKW 接收器的电路理论还不需要波理论，但电缆中能量传播的描述需要经过细化而很复杂的模型。因此人们充分利用在电缆的侧面尺寸小于波长的情况下传播只沿着电缆在一个方向进行的特点。这种单维波理论称为传输线理论，有很重要的意义，因为它还必须用到快速数字电路中。传输线理论概要可以参考参考文献【Her00】、【Hof97】、【Mei92】、【Ste82】以及【Zin90】。

3）高频技术：将频率继续升高，在两个印制电路板之间的连接已经表现出波的传播效应，在一个印制电路板内部甚至生成单个元件时也能觉察到波的传播特性。在这种情况下必须应用波的理论，即应用原始的 Maxwell 公式。计算工作可以利用计算机来完成。理论基础入门可以参考参考文献【Wol97】。

微波技术中各种元器件尺寸的确定也具有很大的挑战性，并且需要很多经验。只有为数不多的工程技术人员从事这项工作。微波技术发展的趋势是购买现成的单元模块，然后将它们集成为一个完整的系统。这些单元模块从系统论的角度来说具有简单可表达的任务，即每个模块只有很少的信号输入和输出端口，如放大器、滤波器、混

频器、振荡器、循环器、加法器、分离器等。复杂的信号处理在低频并且尽可能用数字化的方法来实现，如图 5-29 所示。工程师中的绝大多数只需了解微波元件的黑盒描述，而无需了解它的内部结构。除此之外，测量实践对操作微波元件是很有用的。

5.5.1 天线

使用天线可以发射电磁波，此时要区别近场（离天线几个波长的距离）和远场。技术上有意义并且数学描述上比较简单的是远场。

当场完全与它的源分离时称为辐射。这种情况下源不能确定所辐射的能量在哪里被吸收或者被继续传播。相反，当能量通过感应传输时，负载和发生器是相关的。

要想达到一定的有效系数，天线的尺寸必须要是与波长可相比的。结果如下：

1）天线的计算需要波理论，数学上非常具有挑战性。天线很少从基础做起，多数情况下人们只是从众多产品中选择一个合适的类型。

2）在费用比较合理的前提下，辐射从约为 100kHz 起才是可行的，此时的波长为 3km。

天线要完成两个任务：

1）将导入馈线中的波转换成适合在空间中辐射的波（尽可能少的辐射），或进行相反的工作（接收天线）。

2）将电缆的波阻抗从例如 50Ω 转换到空气（以及真空）的波阻抗 $Z_{W_0}=377\Omega$。

经常讨论的理论上的天线是全向辐射器（点状，没有方向效果的理想天线）以及 Hertz 双极子（电子的短天线，$l\ll\lambda$，沿着天线场强为常数，因此场的计算相对简单）。

所有在实际中使用的天线都是基于 $\lambda/2$ 的双极子构成的。关于它们的形成可以想象一下将长度为 $\lambda/4$ 的双绞线展开，如图 5-42 所示。多数情况下天线的大小为波长的一半。

10MHz 以下的双极子天线可以用线来实现，超过 10MHz 用管或棍来实现。水平悬挂的双极子是水平极化，它有两个主要的辐射方向，横穿过两个极化分支。双极子在导线的方向上没有辐射。

当双极子像图 5-42 右图那样垂直悬挂时，会产生圆形的水平图（圆形辐射器）。在短波波段（波长为 10～100m）通常双极子的一半用管来实现，而另一半用根据位置条件而锚定的线来实现。

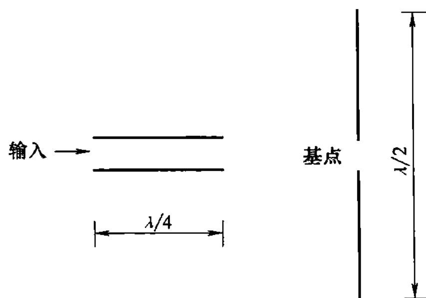


图 5-42 通过展开 $\lambda/4$ 的传输线而形成 $\lambda/2$ 双极子

可以充分利用场论中较熟悉的对称原理，将 $\lambda/4$ 的管埋在（导电的）地下，而将另外一半去掉。这样可以得到地平面天线。VHF 频段内的机动车辆上的天线也是充分利用这种效果（棍状天线）。

天线的最重要的参数包括：

（1）水平和垂直辐射图

1）主瓣的张角；

2）副瓣的最大值；

3）前/后向关系。

（2）与全向辐射器、Hertz 双极子或 λ/2 双极子相比的增益

（3）极化特性

（4）基点阻抗

（5）有效系数

（6）可使用的频率范围

（7）天气阻抗（风力载荷、结冰情况等）

（8）化学阻抗（酸雨的腐蚀）

（9）价格

无源天线具有相反的结构，即方向性、增益、基点阻抗等特性对发送和接收都是适用的。有源天线包括一个具有噪声的预放大器，只能在接收的情况下使用。

接收天线从电磁场接收的功率为 P_E，将其通过天线的导线（通常为同轴电缆）继续送给接收器。下面的关系成立：

$$P_E = \vec{A}_w \cdot \vec{S} \tag{5-37}$$

其中 $\vec{A}_w$ 为有效天线面积，是对一个虚面积归一化的矢量，它的值表示有效系数（增益），方向表示天线的方向效果。$\vec{A}_w$与几何上的面积没有直接的关系，只是由于量纲的原因简单地将它拿过来使用：$\vec{S}$ 的量纲是 W/m^2，P_E的量纲为 W，因此有效天线面积的量纲必须是 m^2。

超过 30MHz 时发送和接收天线必须使用相同的极化方式工作。UKW 无线广播因为机动车天线的缘故使用垂直极化来发射。相反电视使用水平极化，因为潮湿的垂直结构（被雨淋的树木等）对波的吸收较少。当频率在 30MHz 以下时，极化受到电离层的散射会发生旋转，这样在接收端是不同极化的混合。

在自由空间传输时（卫星、定向广播）使用不同的极化可以实现解耦合。这使得更有效地利用无线频谱资源成为可能（极化复用）。

定向天线会影响辐射图（聚光灯原理）因此会产生天线增益。这个增益被定义为与带有参考天线（全向辐射器、Hertz 双极子或 λ/2 双极子）的发射站点相比功率的显著增加。即使用功率为例如 100W 的发射器和一个参考天线，在接收端分别测量信号的大小，然后将发射器换成定向天线，方向对准接收器，再次测量接收信号的大小，并计算以 dB 为单位的放大系数。同样也可以减小发送功率，但在接收端还可以收到与使用参考天线一样大的信号。这个功率减小值同样等于天线增益。

定向天线通常由多个单个天线组合而成。它们的总增益增长很快，这意味着频率越高，由于机械上的原因天线增益越容易实现。

一种应用很广泛的天线是 Yagi 天线，它由多个单元组成：

1）一个 λ/2 双极子作为辐射器；

2）1～3 个加长的 λ/2 双极子作为反射器；

3）多个（1～20 个）$\lambda/2$ 双极子作为控制器。

反射器和控制器只通过场与辐射器进行耦合，因此也被称为寄生单元。它们会通过场产生共振，因此也能将自己的能量辐射出去。所有单元的场相互组合，通过干扰作用会产生希望的方向。图 5-43 所示为一个 4 个单元的可用于例如电视接收的 Yagi 天线。

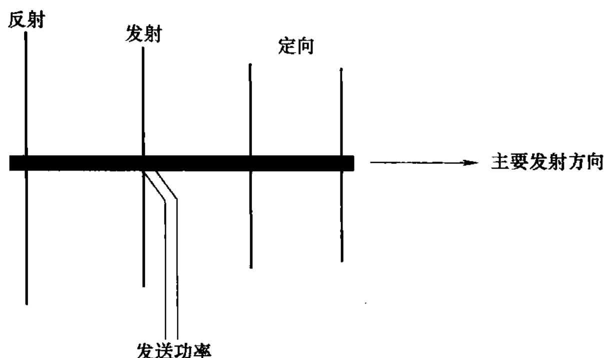


图 5-43 由 4 个单元构成的 Yagi 天线

使用 Yagi 天线可以产生超过 20dB 的增益。除此之外还可以将多个定向天线组合在一起（天线堆、相位阵列、天线组）。如果将一个组的各个天线与可控制的网络联系在一起，那么组的方向特性可以与接收条件相适配（智能天线）。

Yagi 天线被用在 7MHz～1GHz 这个频率范围内。7MHz（波长为 40m）以下天线从机械上来说过于复杂，高于 1GHz 抛物线形天线更有优势。Yagi 天线的带宽相对较窄，因此不同的广播频段（见表 5-2）需要各自独立的天线。

一些广播业务需要带宽较大的天线，因此使用对数周期天线。它的增益比 Yagi 天线的增益小，因此要使用较高的发送功率来补偿。

带宽不仅会影响有效系数，而且也会对基点阻抗产生影响。它们都会随频率发生变化，但馈电电缆的波阻抗应该为常数，这样才能获得波适配及无反射的运行。

当频率超过几百兆 Hz 时，平面天线比较受欢迎。这是由多个被放在一个平面内的单双极子组成的组。它们被用在例如移动电话中。

抛物线形天线的工作原理同汽车的聚光灯是相同的：抛物线形的反射器将辐射汇聚成一个方向的点源。主要的辐射来源于双极子，几 GHz 以上也可以使用法兰状的辐射器，它们被放在抛物线的焦点上。图 5-44 给出了 3 种不同的构成形状。

抛物线形天线的辐射绝不仅仅在单一一个方向上，图 5-44 中左图已经表明了这点。中央馈电的缺点是馈线与主辐射源的连接很复杂。Cassegrain 天线使用一个辅助反射器，从这点上来看是较好的。但辅助反射器会有些遮蔽主反射器，因此会影响天线增益。偏移天线克服了这个缺点。

抛物线形镜面的天线增益与镜面的直径和波长之比有关。正是这个原因汽车聚光灯比抛物线形天线强很多。关于抛物线形天线的增益，镜面直径 d 与频率 f 有下面的近

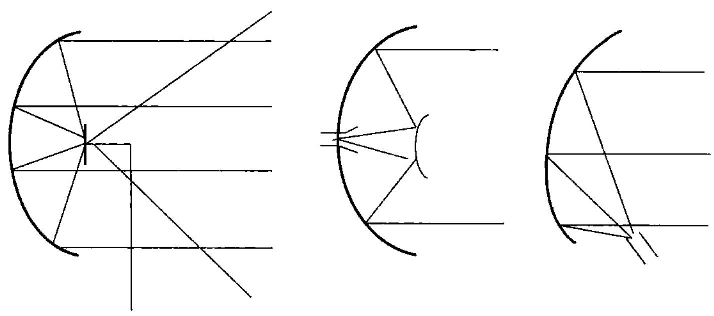


图 5-44 抛物线形天线：左：使用双极子作为主要辐射源的中央馈电抛物线形天线
中：使用法兰作为主要辐射源的 Cassegrain 天线
右：偏移天线（Muschel 天线）

似关系：

$$G[\mathrm{dB}] \approx 20\lg(8d[\mathrm{m}]f[\mathrm{GHz}]) \tag{5-38}$$

作为经验公式的 ±3dB 张角为

$$\vartheta_{\pm 3\mathrm{dB}}[°] \approx \frac{21}{d[\mathrm{m}]f[\mathrm{GHz}]} \tag{5-39}$$

例如直径为 1m 的镜面当频率为 4GHz、张角约为 5°时增益约为 30dB。图 5-45 给出了式（5-38）和式（5-39）的值。

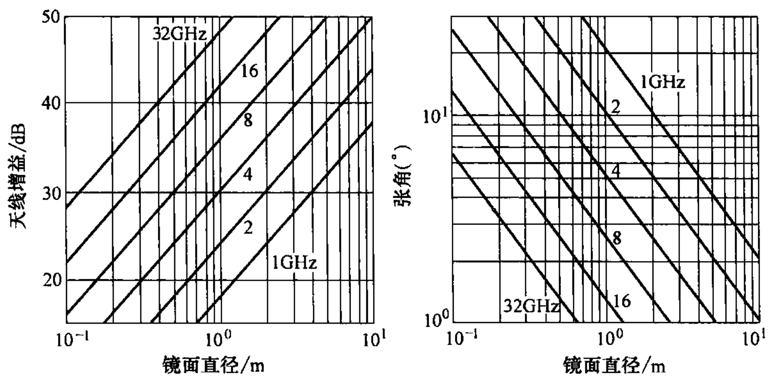


图 5-45 抛物线形天线的增益和张角（它涉及的是参考值，非理想性已经考虑在内）

遗憾的是天线的影响（以及非正确选择和错误操作所造成的损失）经常被低估。一方面，10～60dB 的增益对节省功率以及提高信噪比来说是巨大的，另一方面，定向天线还有助于保护接收器使它不从另一个方向上的不希望的发送器接收数据，这也有助于提高信噪比。

关于天线的其他内容可以参见参考文献【Mei92】和【Zin90】。

5.5.2 传播效果综述

1901 年 12 月 2 日 Marconi 首次用无线电波成功跨越大西洋。Rayleigh 做了一个实验，规模上比跨大西洋的实验要小。困扰他的是 Marconi 的无线电波与 Maxwell 理论相反，并不沿直线传播。Heaviside 和 Kennely 分别假设在空气层之上存在一个反射层，然后解释了 Marconi 的成功。但那时对这个反射层，即所谓的电离层的具体的影响机制还不十分了解。

不受干扰的电磁波会沿直线传播。但干扰或传播路径的不规则性会影响电磁波的传输。影响会分成：衰减效应（吸收）和偏离效应（在电离层的衍射、反射、传输路径上的“坚硬”物体所造成的折射和散射）。

两种效应都与波长和频率有着很大的关系。一般规则是，传输路径上的物体与波长相比越大，物体所造成的衰减就越大。一个长波会绕过一座山脉，相反超短波由于遮蔽作用在山的背面就收听不到了。频率名称和相应的频率值参见 1.1.11 节的表 1-10。

频率越高，越容易出现反射、折射和散射。基于这个原因无线电话工作在 UHF 频段，以便在城市中的街道峡谷中也能工作。

每个发送天线都发送所谓的空间波和地面波。这些名字是根据所辐射的波的跨度给出的。根据频率的不同，空间波和地面波的重要程度不同，参见 5.1.1 节中的表 5-1。为了研究波的传播效果，将频率分成了 4 组：

1）30 ~ 300kHz 的频率（长波）：白天空间波会被电离层吸收，因此在技术上是没有意义的。地面波沿着地球表面传播，传导的地球表面和传导的电离层像一个波导一样共同起作用。使用足够高的发送功率（几百千瓦）也能到达两极（地球上相对的两点，距离为 20000km）。夜晚由于电离层的反射作用传播范围会增加。

2）300kHz ~ 3MHz 的频率（中波）：中波是个混合体，白天的特性像长波，但夜晚又像短波。

3）3 ~ 30MHz 的频率（短波）：地面波很快就被山包等吸收掉（范围为 20 ~ 30km）。技术上没有任何意义。空间波在电离层受到衍射（通常简单地用反射来描述）又返回到地球表面，然后又会被反射，如此进行下去，最后经过之字形的路线以较小的功率（100W）到达地球的两极。但电离层的特性随时间的变化是很强烈的。

4）高于 30MHz 的频率：空间波会穿过电离层而到达宇宙空间。技术上，只有卫星通信对这种空间波感兴趣。地面波的传播几乎像光一样，即带有所提到的散射等副作用沿直线传播。频率越高，传播越像光。由于地球是圆的，它在地面上只能传输较短的距离（30 ~ 100km）。

在 30MHz 以下只能传输窄带信号（大约 10kHz 的带宽），原因如下：

1）带宽根本不是现成的，因为：

LW 和 MW 波段太窄，参见表 5-1 或表 1-10。

在较大的范围内，只允许有少量的资源被同时占用。

2）传播特性与频率有很大的关系，这会导致较大的失真。

超过 30MHz 的频段可以传输宽带信号（几 MHz）。

5.5.3 3MHz 以下频段的传输

总的来说在这个频段内使用窄带信号来工作（音频信号和低速数据传输）。

在 VLF 波段（3～30kHz），发送天线及必要的传输功率的成本对远距离传输来说是巨大的。尽管如此，这个频段也被使用，主要是因为下面两个特殊原因：

1）传播与电离层的影响无关。这主要被 OMEGA 导航系统（12kHz）所利用。这种方法基于对位置已知的不同发射器的信号的干扰测量。它的前提条件是接收信号的相位仅与传输距离有关。基于这个原因人们使用 VLF 来传输字符。

2）VLF 信号可以作为唯一的无线电波穿过水下 10～15m（集肤效应）。这可以使与下沉并因此被掩盖起来的潜艇的通信称为可能。

在 LF 波段（长波，30～300kHz）同样建立起了导航系统（100kHz 上的 LORAN，71kHz 和 86kHz 上的 DECCA）和时间信号发射器。例如在德国使用在 77kHz 上发送的 DCF77 来控制无线钟表。目前 VLF 和 LF 导航已被由卫星支持的更精确的 GPS（global positioning system）系统所代替。

从 150kHz 开始可以使用无线广播。夜晚传播的距离会增加，因此可以减小发送功率。

中波波段（300kHz～3MHz）被船用无线电台和无线广播所使用。

5.5.4 3～30MHz 频段（短波）的传输

最不明显的是适合于窄带信号传输的短波的传播机制。但由于多种原因它们对长距离传输有着特殊的重要性，如图 5-46 所示。这里有 3 条曲线：

1）曲线 a 是不使用中继站点所能传播的范围（参见 5.5.2 节）。

2）曲线 b 给出了大气噪声的值，即在低频部分的值比在高频部分的值大很多的所谓的 $1/f$ 噪声。

3）曲线 c 所示为天线的影响（增益、方向性），参见 5.5.1 节。

图 5-46 中做了标记的范围是远距离传输感兴趣的范围：短波。

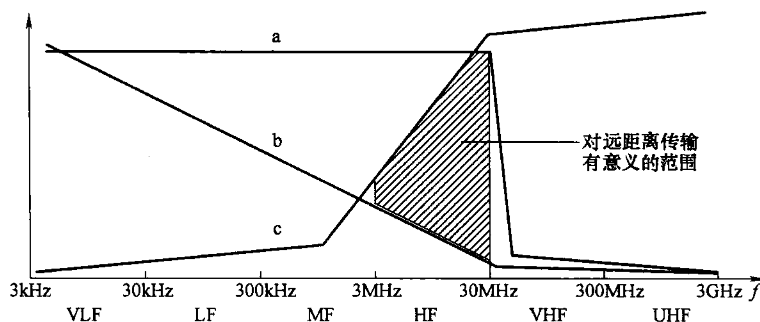


图 5-46 适合于远距离传输的短波所在的特殊位置

短波的最大特点就是电离层对它的影响。电离层就是被离子化的多个空间层，即大气之上的充满导电气体的空间。物理学家将被离子化的气体称为等离子。

电离的机制可以用下面的方法来解释：空气的主要组成成分是氮气（N_2）、氧气（O_2）、氢气（H_2）和二氧化碳（CO_2）。太阳充满能量的紫外线辐射将分子电离，即电子从相应的原子中分离出来并自由运动。如果它们落到带有正电的原子核附近，原子核就又将它们捕获（重组）。电离的速度与太阳辐射的强度有关，重组的速度与分子的密度，即在一定的时间内遇到一个分子的概率有关。

自由电子的密度，即每立方米自由电子的数量，随高度变化。高于 400km 紫外线辐射特别强，但分子密度很小，因此自由电子密度也很小。越往下分子的密度越大，因此电子密度也会升高。继续往下，紫外线辐射变得很弱，但因为引力的作用分子密度很大，重组的速率大于电离的速率，因此电子的密度会下降。电子密度在一特定的高度达到最大值，这就构成了所谓的电离层。

太阳的紫外线含有不同波长的波。空气中的不同气体对这些频率的反应不同，因此随着电子密度的升高构成了多个层次。这些层次用多个大写字母来描述，参见表 5-7 和图 5-47。描述从 D 开始，因为那时的研究人员还不知道，是否在 D 层之下还会继续发现新的层次。

表 5-7 电离层的层次

层次	高度/km	遇到时间	影响
D	50 ~ 90	只有白天，最大值在本地时间 12:00	对长波造成反射 对短波造成衰减
E	90 ~ 125	只有白天，最大值在本地时间 12:00	对较长的短波造成反射 对较短的短波造成衰减
F_1	150 ~ 250	只有白天，夜晚与 F_2 层组合成 F 层	对较短的短波造成反射
F_2	350 ~ 500	只有白天，夜晚与 F_1 层组合成 F 层	对较短的短波造成反射

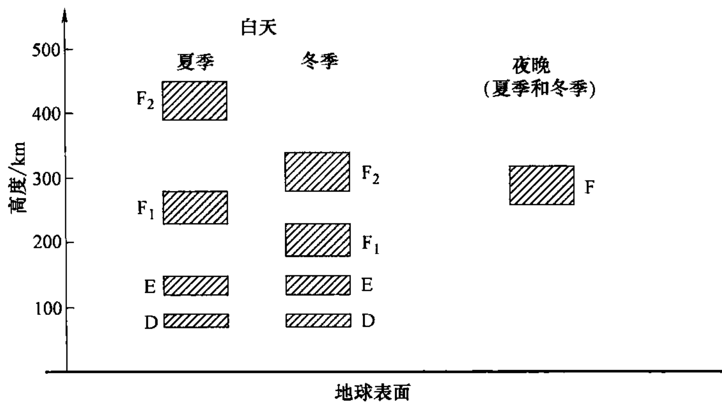


图 5-47 电离层的层次

对短波传输来说，最重要的层次是F层（包括两个子层次）。由于在较高的地方密度较小，因此重组的速率很小，以至于F层只在夜晚才能出现。高度的比较：天气只在10km以下才起作用，宇宙空间站（ISS）在距地球表面350～400km处飞行。

D层是很讨厌的，因为它会造成衰减。D层在空中是散射的，为此纯化论者说是D区而不是D层。

因为电离层与太阳的辐射有关，它在时间上会有波动：会随一天中的时间和季节发生变化。

人们已经确定，太阳紫外线辐射的强度并不是常数。实际上它与太阳黑子（即太阳表面较黑的区域）的数量有关。太阳黑子的数量有着11.6年的波动周期，原因还不是很清楚。除此之外还有27天的波动，这是由太阳本身的自转引起的。太阳黑子会被成组的遇到，并且太阳黑子在偏离地球的一边对电离层不会造成影响。经过整个的旋转，会遇到只存在有一个黑子组这样的机会。

顺便说一下，天文学家将我们的太阳看成是所谓的G类矮星的非常不引人注意的一个代表。这一类中的所有恒星好像都有太阳黑子。对周期的测量明显地显示是到现在还无法解释的约11年的时间，很少会出现低于3年或高于30年的例外。经过一个太阳黑子周期，太阳磁场的极性会发生变化。

一对太阳黑子构成磁力线的出口点和入口点，磁力线从太阳出发又回到太阳，形成环状，就像杯子的手柄一样。太阳黑子从太阳内部得到少于太阳其他表面的能量，因此它们的温度约为4500℃，看起来比它周围部分的约6000℃的地方要黑。

2000年前中国的天文学家就观察到太阳黑子，1611年伽利略又发现了它们。苏黎世ETH大学的天文台从1849年就开始研究太阳活动，对从1750年开始的记录进行了重新评估。根据苏黎世太阳观察点的创始人Rudolf Wolf的建议，通过计算太阳黑子组g和单个太阳黑子的数量f，并因此规定了苏黎世相对值R，对太阳的图像进行了重新解释。

$$R = k(10g + f) \tag{5-40}$$

当然这里R具有主观的色彩，所使用的光学仪器（影响式（5-40）中的系数k）以及所评估的值的数量对结果都会造成影响。

太阳黑子爆发时，以宽带电磁辐射的形式喷发巨大的能量。电磁辐射从无线广播的频率范围，经过紫外线一直到达X射线。黑子较大的辐射对短波的传播有巨大的影响。根据能够被很好地观测到的太阳黑子可以间接地得到辐射的量化值。但是R每个月的波动比黑子的波动大很多，因此人们用一个平滑的中间值计算经过平滑的苏黎世相对值R_S（S代表平滑）。图5-48所示为1750年到1998年的R_S值。根据这张图可以很好地识别出太阳黑子数量的周期。从1755年开始对这些周期进行了编号，第23个周期开始于1996年底。图5-49是对第21和第22两个周期R和R_S的比较。

R_S构成了对无线传播进行预测的基础。

当然人们一直在寻找对太阳活动进行测量的客观方法。为此，通过在波长为10.7cm（频率为2.8GHz）时未得到太阳辐射的强度的办法来直接定义黑子的影响。这个太阳通量（Solar Flux，SF）与经过平均的太阳黑子相对数量R_S有非常密切的关系。

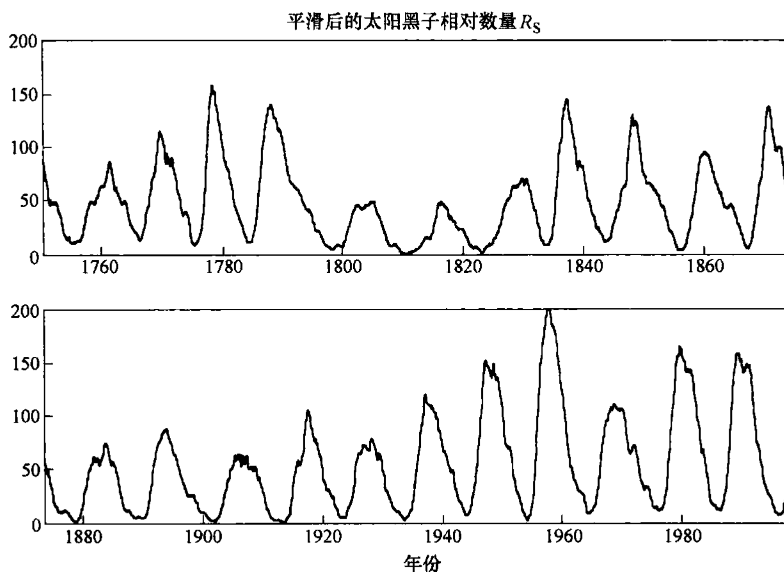


图 5-48　从 1750 年到 1998 年平滑的苏黎世相对值 R_S

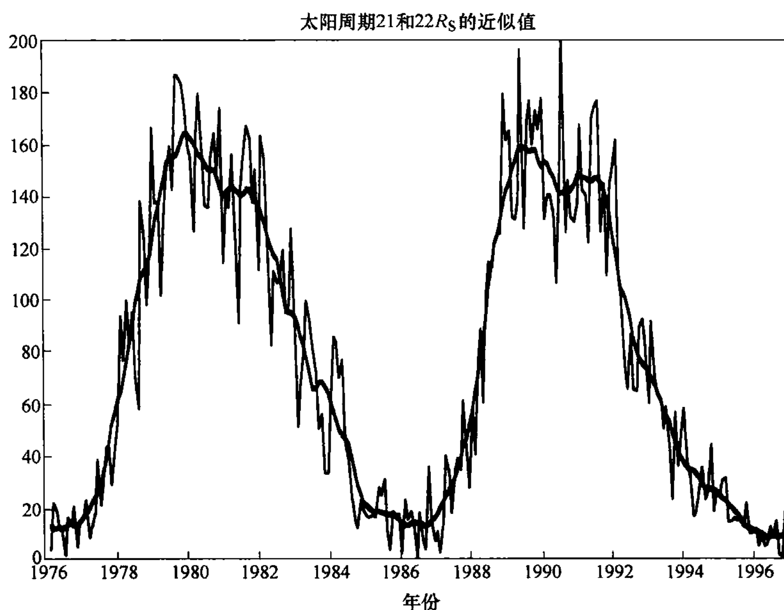


图 5-49　第 21 和第 22 周期 R 和 R_S的比较

这种关系可用幂级数来表示，但这里二次方项就已经很小了：

$$SF = 63.75 + 0.728R_S + 0.00089R_S^2 + \cdots$$

在实际中用线性方法来换算就足够了：

$$SF \approx 63.75 + 0.728 R_S \tag{5-41}$$

这里能体现出当时 Wolf 提出式（5-40）是多具有独创性，或至少是种直觉。太阳通量每天都会被不同的天文台给出，实际结果很容易例如通过互联网来获得。

R_S和 SF 描述了电离层电离的原因，即太阳活动。另一种方法是直接测量电离作用本身的影响。电离会通过电离层反射的能力来间接地得到。为此使用一个音响器将频率交替变化的信号垂直向上辐射，然后用所谓的离子图对回波进行评估。这同回声探测器差不多，只不过是用电磁波代替声波，并且向上辐射而不是向下辐射。刚好能够被反射回来的频率称为临界频率f_0。每一层都有它自己的临界频率，用f_{0E}、f_{0F1}等来描述。世界范围内共安装了 150 个这样的音响器，图 5-50 给出了一些测量结果。

短波波段的较低频率会被 D 层给吸收掉。升高频率，衰减会变小，可以观察到被 E 层的反射。继续升高频率，那么会穿过 E 层，又能观察到从 F_1 层的反射等。F_1 层也能反射较低的频率，这是因为位于它下面的各层根本不够厚。

根据回声的运行时间可以计算出反射层的虚高度。事实上这里会发生衍射，但计算上可以将它作为反射，换算成一个虚高度。

图 5-50 中的非正常的辐射是由双折射而引起的，主要是因为地球磁场对电离层所造成的影响。双折射跟在一些水晶上产生的光学上的双折射类似。经过电离层辐射会变成两个，这两个辐射由于虚高度不同会有不同的传播时间。这样会以多径接收的方式产生线性失真。非正常辐射在 F 层和 F_2层是最强的，F_1 层稍弱，E 层几乎很难测量到。

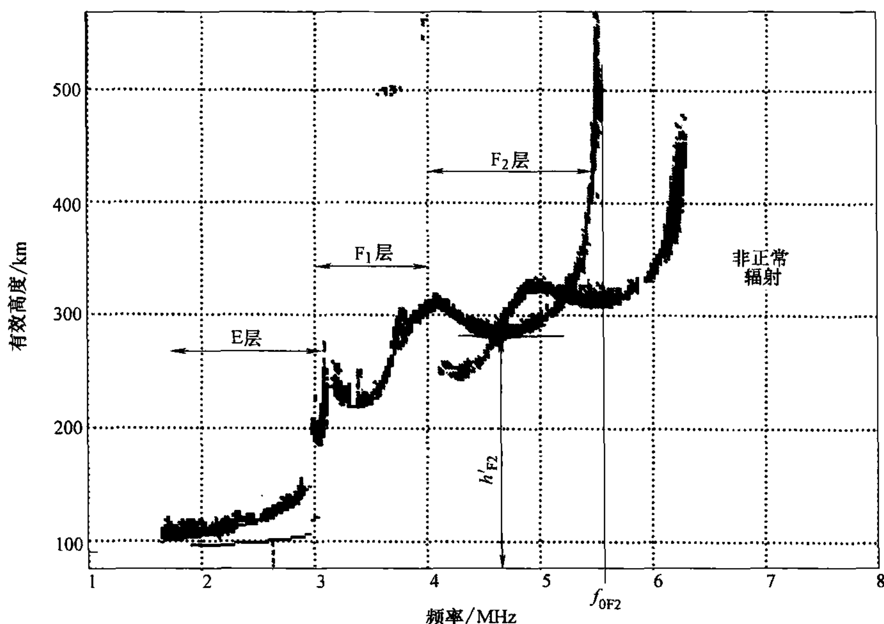


图 5-50　对部分根据带有解释的离子图（实际测量）

图 5-51 所示为临界频率随一天中的时间和 R_s 的不同值而变化的关系。图 5-52 是临界频率与地理纬度的关系。地理和时间的波动与太阳的照射时间有关，而 R_s 值能刻画太阳辐射的强度。

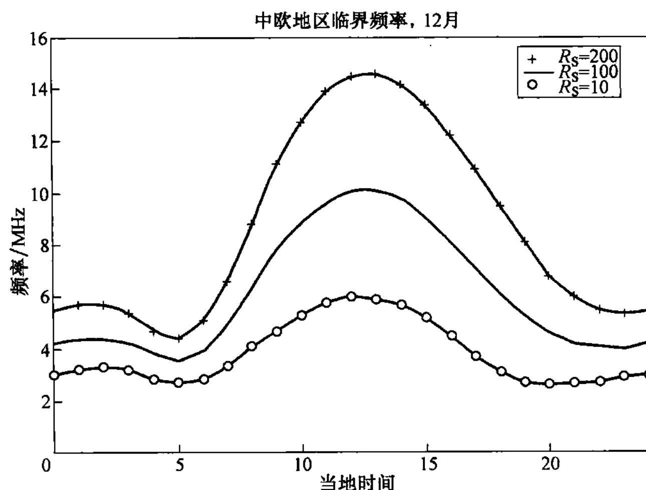


图 5-51　临界频率与一天中的时间和太阳活动的关系

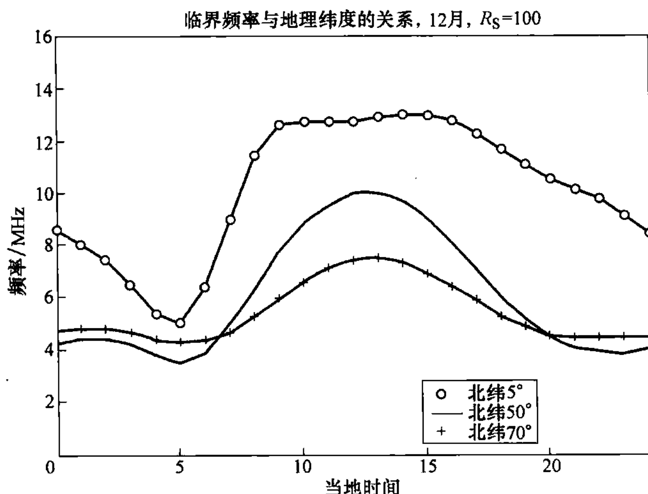


图 5-52　临界频率与地理纬度的关系，此时经度为常数，并且在同一个月份。

临界频率（物理学家称为等离子频率）是当辐射垂直进入电离层时刚好能被反射回来的最大频率。在长距离传输时，传输路径是围绕地球沿“之”字形进行的。借助于 F_2层的一跳（英文为 hop）距离约为 4000km。人们试图采用能使发送天线中的辐射尽可能以水平的方式来工作，以便总的跳数能尽可能地少。这样穿过 D 层的辐射数量也会减少，因而衰减也会变小。

水平进入电离层的辐射，当超出临界频率时，在一定的极限内也会被反射。此时人们用最高可用频率（Maximum Usable Frequency，MUF）来代替临界频率。MUF 可以在已知发送器和接收器地点的情况下，根据纯地理高度 h'和反射点的临界频率来计算。后者可以用一个垂直音响器（必须在反射点下方，不一定总是可能的）来确定，或用一个倾斜的音响器（带有一定角度的辐射器，直接遗漏 MUF），或根据其他地点的已知临界频率使用插值和预测的方法得到。

这样计算得到的 MUF，因为有插值的存在是不精确的。除此之外人们在计划未来的连接时并不知道 R_S的精确值，只有它的预测值。因此人们并不使用 MUF 的结果，而是使用最优传输频率（Frequency of Optimal Traffic，FOT）。它的值是 MUF 的 85%，因此拥有一定的安全范围。MUF 和 FOT 是与发送功率无关的。此外还需考虑最低可用频率（Lowest Usable Frequency，LUF）。LUF 描述的是连 D 层都能穿过的最低频率。因为它涉及到的是吸收，所以 LUF 与发送功率有关。从定性的角度来说，LUF 白天像临界频率。中午 LUF 的值最高，晚上它下降到短波范围。

在较长的白天时段，LUF 的值甚至会超过 MUF，因此在这个时段 KW 连接是不可能的，即工作频率要高于 LUF 而低于 MUF。在长距离传输中多跳是必需的，在这种情况下，较低的 MUF 对所有的跳数都是适合的，而衰减要逐段相加。

小结：对短波的传播起关键作用的是电离层电离的强度。电离与太阳辐射的强度和时间有关，它会有下面的波动（已经标在图 1-16 和图 1-17 中）：

1）随季节变化的波动；

2）太阳黑子的周期（周期为 11.6 年）；

3）太阳的自转（周期为 27 天）。

为测量离子的数量，下面的值起主要作用：平滑的苏黎世相对值 R_S，或者太阳通量。

从 R_S 或 SF 可以推导出或可以直接测量到的起次要作用的值为临界频率 f_0/离子图。

临界频率描述在一定的位置、一定的时间上电离层的状态。对无线连接来说，人们还指定发送器和接收器的地理位置。

为此，对特定的无线连接，根据 R_S 或 SF 可以计算：MUF（由于衍射而造成的）、FOT（MUF 的 85%）、LUF（由于衰减而造成的）。

趋势：

（1）电离的强度

1）白天电离的强度要大于夜晚。

2）夏天电离的强度要大于冬天。

3）太阳黑子活动最剧烈时电离的强度要大于太阳黑子活动最弱时。

（2）影响：电离增强时

1）MUF 会增加，因为较高的频率也会被 F 层反射回来。

2）LUF 会增加，因为 D 层的衰减会增加，经验值：频率值增加一倍，衰减值会减少到原来的四分之一；发送功率增加 10 倍，LUF 会下降 2MHz。

(3) 对短波通信时频率选择的结论：

1) 电离较强时比电离较弱时使用的频率较高。

2) 白天时，由于 D 层的衰减，小于 5MHz 的频率几乎用不到。相反夜晚，这些频率能够被很好地利用。

3) 大于 20MHz 的频率主要在太阳黑子活动最剧烈时或白天使用，因为 5W 的发送功率对洲际间的连接就够用了。而相反当太阳黑子活动较小时，这些频率只能是保持无线电寂静。

4) 短波通信时人们一直在试图使用尽可能高的频率，这样 D 层的衰减会最小。但频率也不允许太高，否则在 F 层就不会有反射。

5) 只有当频率在一天当中变化多次时，使用短波的无线连接才可能永久地保持下去。

由于在一天中的波动，在某一特定的时间只有一部分短波范围是可用的。在这种情况下，所有的用户争相使用这些频率，使得这些频率间相互出现干扰。干扰的主要部分不是噪声，而是人为的、以很强的窄带信号形式出现的噪音。在这种情况下，在站点旁需要一个有经验的、耐心的并且有直觉的操作员。

因为 MUF 和 LUF 一直在变，人们对它们的实时值并不了解，人们使用预测的 R_S 值来工作，即无线广播的天气预报。当然这种预测有一定的不安全性：使用预测的 MUF 工作，通信成功的概率为 50%，而使用预测的 FOT，概率可达 90%。

未来通信的频率规划（计算 MUF/FOT 和 LUF）需要下列输入：在一天中的时间（解析度：1 小时）、季节（解析度：1 个月）、R_S或 SF 的预测值、发送器和接收器的位置。

场强的预测还需要发送功率以及发送天线和接收天线的增益。

以前人们使用大量的表格来预测无线连接，现在有舒适、便宜的 PC 程序。R_S或 SF 的预测值可从小册子或互联网上得到，不同的天文观察站都公布实时的测量值。

当前传播状况的实时图也可以通过自己的测量来得到。为此在世界范围内广播并且同步一个无线电导航台（beacon），它们是功率和无线发射天线都是已知的、全自动的工作发送器。如果能听到导航台，MUF 就在更高的地方。

图 5-53 给出的是一个从中欧到复活节岛（南太平洋）这一段无线电波预测的例子。时间使用的是 UTC（Universal Time Coordinated），比 MEZ[⊖] 时间晚一小时，并且不是夏令时（以前 UTC 被称为格林威治时间，即 GMT）。

图的上部分给出了两个端点在一天中的时间，它的下面就是 MUF 和 LUF。很典型地，当两个地方都是夜晚时，LUF 的值下降，因为此时离子数也在下降。UTC 的 20 点之后，MUF 下降到 LUF 处，在这段时间内，在实际中短波通信是不可能的。

图的下部分描述的是几个频率的接收情况，它们都是业余无线电爱好者使用的频段。只有当 MUF 足够高时，28MHz 的那条线才会出现，相反只要 LUF 升高，3.5MHz 和 7MHz 那两条线就消失了。有代表性的是，7MHz 那条线包围了 3.5MHz 那条线，因为两者主要都与 LUF 相关。

⊖ MEZ 指欧洲中部时间。——译者注

无线电波将两点在地球表面上沿着大圆弧连在一起。这个圆弧位于通过发送地、接收地、地球中心点扩展后所形成的平面上。这个大圆弧被分成两部分，即短路径（short path，始终小于 20000km）和长路径（long path，40000km 减去短路径）。对距离大于 10000km 的连接而言，在以下两种情况下选择长路径有优势：

1）频率低于 10MHz，长路径经过地球上的夜晚时，此时可以充分利用 LUF 较低的优势。

2）频率高于 10MHz，长路径经过地球上的白昼时，此时可以充分利用 MUF 较高的优势。

前提条件当然是可用可旋转的定向天线，这种情况在频率高于 10MHz 时是很常见的。长路径和短路径的区别是方向相差刚好 180°。图 5-54 所示为使用长路径连接复活节岛的预测，此时使用的参数与图 5-53 中所给出的是相同的。在 UTC 14 点时，只有通过长路径与复活节岛的 KW 连接才是可能的。

还可以对参数 R_S（例如 10、30、60、90、120、150）、一天中的时间（例如所有小时）和季节（例如所有月）进行详细的描述。为此需要计算 $6 \times 24 \times 12 = 1728$ 个世界地图的地区图，用不同的颜色给出 MUF 的值。图 5-55 所示为一个实例。

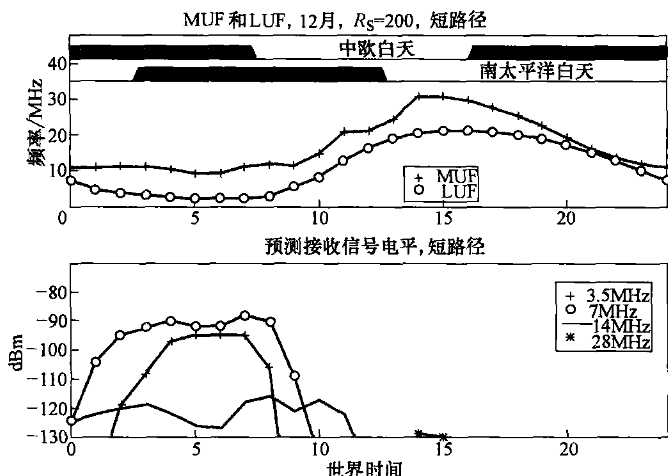


图 5-53 从中欧到南太平洋的复活节岛短波通信的 24 小时无线电波预测，距离为 14000km（计算的前提条件为：12 月，$R_S = 200$，发送器采用 500W 的双极子天线，接收器也使用双极子天线）

现在要详细地解释一下图 5-55。很明显可以看出，南美洲的方向 MUF 比亚洲或欧洲要高。这根据太阳的位置很容易解释。

北非在 20MHz 处经过一跳就能到达 F 层。沙哈拉地区颜色很浅，MUF 的值较低。这个地区一跳达不到，但二跳又太近。因此电磁波必须要稍微倾斜一点地发射，以便每一跳的距离会减小。但对于倾斜发射的电磁波，MUF 的值又会下降。

再往南，MUF 的值又升高。这些地区使用较平的发射角度在 20MHz 的频率上二跳就能够到达。

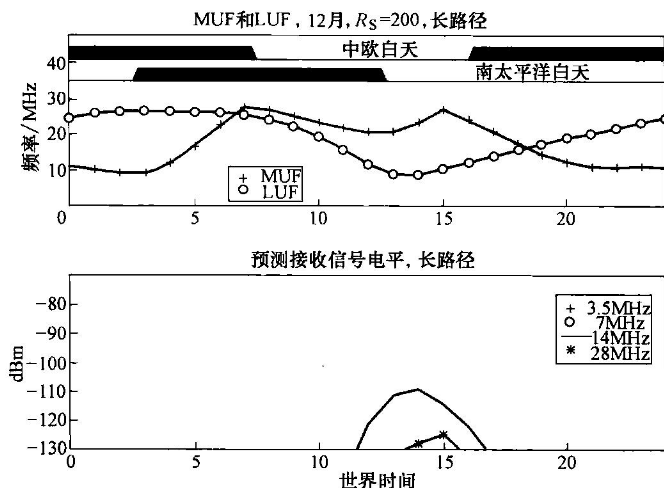


图 5-54　同图 5-53（但计算的是长路径，距离为 26000km）

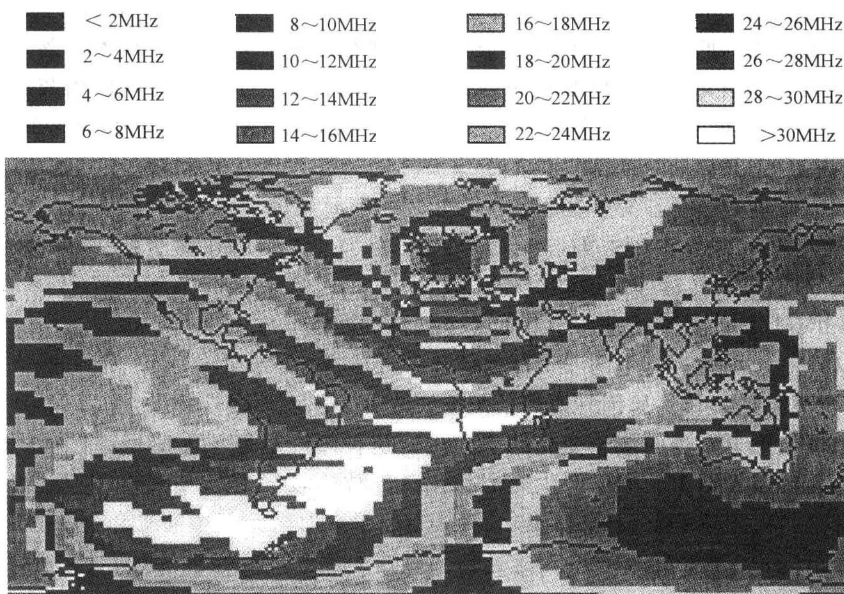


图 5-55　对中欧 MUF 的详细预测（针对 4 月，18：00UTC，R_S = 90 计算）

对更远的距离，到达这些地区的跳数在 n 跳和 $n+1$ 跳之间。电磁波可通过不同的传播路径到达那里，这样会导致干扰（fading）。

从中欧向北，在 20MHz 频率上一跳就能到达 F 层。在北极点地区由于太阳的辐射较弱，对 20MHz 频率上的第二跳而言 MUF 太低，如图 5-52 所示。当 R_S 较大时在中欧周围有第二个圆。这种围绕发射器的同心圆现象被称为 MUF 锥形。

短波传播时会遇到几个令人吃惊的情况。但它们是可以解释的，并可以在做规划时进行考虑：

1）死区：在无线技术的开始阶段，人们很快就发现短波波段的地面波的范围非常受限。而对于空间波，当工作频率小于临界频率时，由于射向电离层的角度很小，所以又回到了发射器附近。但频率经常是较低的，这样通过D层所造成的衰减会使空间波的传输不可能进行。升高频率，电磁波就会进入所有的电离层，没有反射存在。

离发射器稍远些，就会进入死区，这里接收不到任何信号，因为地面波被吸收掉，而空间波还不存在。

离发射器再远些，突然空间波被反射回来。此时进入电离层较浅的电磁波被反射回来。它们的频率小于MUF，但高于临界频率。

由于死区，以前整个短波波段都是业余无线电爱好者的领地。这对短波的传播起了很大的作用。当短波对远距离传播具有很好的使用价值被发现后，在这个频段立刻有了许多商业和国家的用户。业余无线电爱好者在这个频段内只剩下了1.8MHz、3.5MHz、7MHz、10MHz、14MHz、18MHz、21MHz、24MHz和28MHz可用。使用这些频率人们可以在任何时间进行跨洲际的连接，在太阳黑子活动最弱时，使用14MHz也可以到达两极。无数的业余无线电爱好者可以做各种各样无需任何研究院所来付费的实验。

2）灰线连接：灰线是模糊区，这里D层还不能（早晨）或已经不再能（晚上）遇到了。因此特别低的频率的衰减也很小。例如在圣诞节期间，中欧太阳落下的时间与加利福尼亚太阳升起的时间相同。因此这两个区域在这个时刻可以使用灰线进行连接。在20~30分钟的时间内使用较大的信号强度在3MHz上进行通信是可行的。

到现在为止我们讨论的都只是关于短波之间的关系的计算问题。这里还有一些不容易预测的影响，甚至还有一些到目前为止还不能解释的现象：

1）已经提到过的由于太阳自转而产生的27天的太阳活动周期所引起的波动。使用精确的R_S值很少能估计出这种影响。

2）由不规则的太阳活动所引起的完全偶然的波动。

由这些不规则性所产生的非正常的传播效果包括如下：

1）电离层的前移：原因是所谓的闪耀，它是太阳表面强烈的爆炸，怀疑是由较大的太阳黑子组附近磁场的快速变化所引起的。这种爆炸从地球上使用望远镜就能观察到。这些闪耀使太空中的物质产生离心作用，并放出较强的X射线和UV射线。因为这些闪耀是与太阳黑子有关的，所以在太阳黑子最多时遇到闪耀的机会会增加。

闪耀出现8分钟之后辐射会到达地球。电离层引起的衰减会增加，18~36小时后会遇到物质微粒。它们是带电的（质子或电子），会与地球磁场或电离层发生反应。结果是会出现极光线现象（北极光），以及可持续3天之久的电离层磁场中心前移。噪声电平会增加，无线电波会变差。经过两极的路径比穿过赤道的连接所受的影响大。

测量太阳活动的天文台也研究地球的磁场。地球磁场对闪耀的反应（用所谓的K指数给出）以及地球磁场的预测值（所谓的A指数）已经与太阳活动的测量值一起公开发表了。

2）Mögel-Dellinger效果：它是极强的闪耀所产生的结果。几分钟之内衰减增长的特别强烈，以至于会中断短波通信。它会产生很令人吃惊的效果，人们常常会以为是自己的接收设备被损坏了。这种效果会持续15分钟到几个小时。当然它只会出现在地

球上为白天的那一侧。晚一些时候还会引起很强的电离层风暴。Mögel-Dellinger 效果也被称为 SID（Sudden Ionosheric Disturbance）。

3）衰减：有时不仅仅是接收器的几个波，而是同时有多个波都会受到影响。这些波经过不同的路径、不同的运行时间和衰减，例如有 n 及 $n+1$ 跳。在接收地，根据相位的不同，会有加性的或损坏性的干扰。当传播关系发生变化时，衰减会产生所谓的信号强度波动。常见的原因是电离层的电子与地球磁场的相互作用。因此，进入到电离层的短波会被分成两个频率不相同的部分（正常和非常正常的电磁波）。然后这两个波会经过不同的反射和衰减。这在离子图（见图 5-50）中经过 F 层产生的回声的那一栏可以看到。除此之外，电磁波的磁极也会发生旋转。因此在短波中与 VHF 和 UHF 通信不同，不要求发送天线和接收天线有相同的磁极。

不想继续讨论的是短波的行为是非常特殊的，为了找到不断变化的最优频率需要很多经验。因此短波也得到了“是不可靠的传输媒体”的名声，除此之外信道能力也很缺乏。由于在 20 世纪 70 年代卫星技术发展的很快，并且在 1976 年还经历了一次太阳黑子活动最弱期，因此短波在商用方面宣布结束。人们用非常可靠的卫星来完全代替了它，短波技术在实际中的改进工作正式开始。

但这里所说的情况又是另外一个样子：短波经历了复兴。这有多个原因：

1）如果使用卫星通信，那么通信还取决于第三方。并不是所有的通信用户都希望这样。短波本身也能在自己的管理范围内提供世界范围内的远距离通信。

2）短波的主要缺点是它的快速多变的特性。为了克服这个缺点，以前人们将电台做得很好。现在，由于有了微处理器，机器本身也具有可适配的能力。这样现代的微波电台的操作以及基于短波的业务的开发变得与使用电话网来发送传真一样容易。

短波遗留的一个缺点就是：与卫星技术相比，它的信道能力较小。由于存在多径传输和频率选择性衰减，波特率不能超过例如 200 波特，否则会出现码间干扰（ISI）。在短波上只能使用窄带信号。使用多值调制的情况下，有用数据传输速率可以达到约 1～2.4kBit/s，这对电报来说是足够了。话音传输会使用 SSB。现在短波正好还提供这两项业务，例如外交代理、船用无线电台、飞行用无线电台，当然还有军事上的应用。

短波信道可能是所有信道中最难的了，针对较低的 1～2.4kBit/s 数据传输速率，所有的技术手段都得用到。但令人着迷的是这个事实，短波的优良特性是由于偶然才被发现的，然后根据经验进行了研究，最后才变成科学研究的主题。尽管并不是所有的观察现象都得到了解释，在很长的一段时间内短波在技术上得到了很大的应用。

使用短波进行数据传输已经有了很长的时间。最初使用的是摩尔斯电码，目前只有在业余无线电爱好者中还保留着它。

不久人们就开始使用无线电传，使用五位的 Baudot 码实现异步传输，如图 2-1 所示。波特率是 50 波特。调制技术使用的是 170～850Hz 的 2FSK。这种 RTTY 方法直到 20 世纪 80 年代初都是使用最多的方法。

随着微处理器的出现，使用智能方法变得很实际。SITOR（Simplex Teletype Over Radio）是个开始。它已经在 4.3.1 节提到的 ARQ 方法中提供信道编码。

随着便宜和功能强大的信号处理器的推广，情况发生了迅速的改变。现代短波接收器由图 5-56 所示的单元组成。

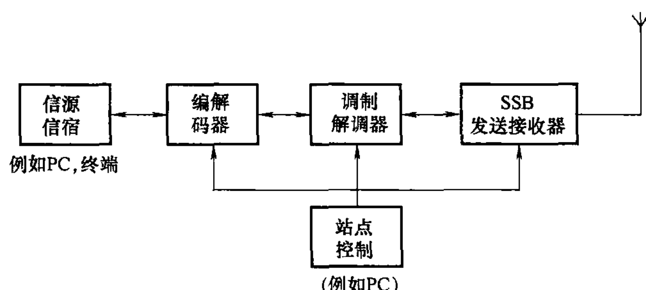


图 5-56 用于短波数据传输的站点的组成

编解码器执行信源和信道的编码工作并组成传输帧。可以有多种选择方案（冗余的多/少、帧的长短）。调制解调器可以处理多种从慢速 2FSK 到快速 16QAM 的调制方式，它工作在音频频域内，用数字信号处理的方法实现。SSB 发送器实现到短波波段的转换工作。为此可使用一个传统的 SSB 设备，但它必须能提供一个较好的频率常数。这个频率最好能用远程可控的频率合成器（参见 5.4 节）来提供。ARQ 希望有一个快速的发送-接收转换装置。对话音和数据传输来说，发送功率在 100W ~ 2.5kW 之间。无线广播的发送功率高于 100kW。

编解码器和调制类型的选择在运行期间是可调节的。连接的建立使用的是慢速并且可靠的方法。如果在输出过程中遇到的差错较少，那么就可以提高数据传输速率。当连接被中断时，站点会自己尝试重新建立连接。这里它会避免使用替代频率。为了达到最优的接收效果，每个站点都能调节对方站点的发送功率。

站点控制器执行较少的实时任务，它们通常是用 PC 来实现的，这也完全可以同时作为终端设备（信源和信宿）。

两个用户之间连接的建立可以采用下面的方法来进行：每个 PC 每小时计算一次传播预测，之后站点就知道了 FOT。因为陌生的站点可能会在相同的区域内也计算出一个 FOT，这样由于频率被占用而引起干扰的危险性会很大。因此每个站点在当前的 FOT 范围内计算出例如 10 个不同的频率。

然后使用一个随机序列样本对这些频率进行持续控制，这样会发现一个可能的呼叫。如果要想发送数据，站点就从这 10 个可能的频率中选择一个空闲的，与对方站点开始呼叫。如果在一定的时间内没有收到应答，站点会使用另外一个频率重试。以前这种单调乏味的工作都是由无线电话务员来从事的，现在可由站点自己来完成，即 ALE（Automatic Link Establishment）。

如果在一个交换中心有多个发送器和接收器在工作，那么信号的接收可能会受到其他连接的较强的发送信号的干扰。在这种情况下，会将发送器放到与接收器相距几千米之外，并进行远端控制。

到目前为止用于数据传输的短波使用的一直都是从 0.5 ~ 1kHz 带宽的窄带信号。

目前研究领域正在使用一种完全不同的解决概念，即扩频（参见 5.3 节）。后续的工作目标是使用短波来传输经过 LPC 编码的话音信号。OFDM 调制（参见 3.4.6 节）由于有较长的符号持续时间，同样不久也能在用于短波上。

最后，我们计算一下在短波波段内非常强的噪声。这些计算基于式（5-7），用天线噪声系数 F_{ant} 来描述。F_{ant} 随着频率的升高而下降，并且与天气和季节也有关系。在短波范围内 F_{ant} 的值在 0 ~ 50dB 之间。有雷电时这个值会明显的受损。

根据式（5-8），接收器的总噪声为

$$P_N[\text{dBm}] = -174\text{dBm} + 10\lg(B[\text{Hz}]) + F[\text{dB}]$$

现在还要加上天线的噪声系数 F_{ant}，这样总噪声增大为

$$P_N[\text{dBm}] = -174\text{dBm} + 10\lg(B[\text{Hz}]) + F(\text{dB}) + F_{ant}[\text{dB}] \tag{5-42}$$

作为数值上的例子我们可以观察一下使用 2.7kHz 的 SSB 通信。接收器的噪声值为 $F = 10$dB，天线的噪声为 $F_{ant} = 20$dB。根据式（5-42），噪声电平为 -110dBm。SSB 通信需要的最小信噪比为 15dB，这样话音才是可理解的（参见表 1-8）。这样接收器输入端的信号电平最小必须是 -95dBm。这相当于大约 0.32pW 的功率或 50Ω 情况下 4μV 的电压值。

从图 5-54 可以看出，对到达复活节岛的长路径而言，需要的电平位于下极限值处。如果使用两个增益值都为 10dB 的 Yagi 天线来代替双极子天线的工作，那么信号电平值就会增大 20dB，通信会毫无问题地进行下去。当大气噪声较小时使用双极子天线就足够了。

5.5.5　30MHz ~ 1GHz 频段的传输

正像已经提到过的，VHF/UHF 的特性已经类似于光。因此对于远距离通信人们通常选择较高的地点发射，以克服地球的曲率的影响。相反，在远距离的短波通信中人们选择较平的辐射，即发射地点选择充分平坦的水平面。高度本身是无所谓的。

总的来说 VHF/UHF 通信是稳定的和可靠的，因此可以使用转接站。转接站位于山峰上或是太空中（卫星）。

由于最大可达 3km 高度的低层大气层的反射，VHF 和 UHF 波可传播的距离比可见地平线还要远些。人们称之为无线电地平线。

VHF 波段（30 ~ 300MHz）同样会遇到传播变异，下面解释一下其中的几个：

1）经过 F_2 层的 VHF 传播：在较强的太阳黑子活动最剧烈的期间，洲际间通信的 MUF 可以升高到超过 50MHz。原则上这是很正常的，但由于它很少出现，所以属于耸人听闻的情况。

2）稀疏 E：偶然并且是很不规则地直接在 E 层下会形成很强的直径为 80 ~ 160km 的电离"云"。这片区域会在附近漂移，但生存期只有几小时。它们对直到 150MHz 的频率都能反射。稀疏 E 在夏天（主要是在当地时间 10 ~ 12 点和 18 ~ 20 点）会比在冬天（主要是在当地时间 21 ~ 23 点）遇到的更频繁。这种不能解释的现象在 TV 远程接收时也会遇到。

3）极光通信：北极光的起源我们已经解释过了。如果太阳的粒子流足够强，那么

当短波的衰减增加时，极光就能反射 VHF 无线电波。由于这些衰减（100～1000Hz 上衰减的波动），话音信号很勉强地才能够被理解。但摩尔斯电报的通信是可行的。

4）流星散射：在大气层的上层（大约 100km）燃烧掉的流星（消失的恒星）留下已经电离的尾状物，它能够反射 VHF 信号。这种现象会持续几秒到半分钟的时间，允许直到 2000km 的通信，这个距离是由流星尾状物的高度和只有一跳的事实来决定的。每天会有上亿颗流星（其中绝大多数的质量都在毫克的范围，其实是灰尘微粒）燃烧掉，因此可以从技术上充分利用这种效果。如果地球轨道与流行雨轨道交叉，经常会有连接的可能性出现，例如在 8 月 12 日的 8 月流星群出现时。

发送器向接收器发送持续几毫秒的突发类型的呼叫，其中的绝大多数都没有成功。如果遇到一种连接可能性，那么接收器就会应答，发送器以几 kBit/s 的速率发送信息。较长的信息会被分成包单独进行传输。平均每天有用数据的吞吐量约为 50～75Bit/s，这相当于一个永久存在的电传信道。

在一个测试中，要测量一条被分成 8 个分组的消息的传输时间。结果是很令人吃惊的：约 40% 的情况下需要的时间少于 30s，约 80% 的情况下需要的时间少于 90s，约 95% 的情况下需要的时间少于 200s，没有超过 400s 的情况。

因为传输距离在空间上是很有限的，并且在时间上是偶然的，MBC（Meteor Burst Communication）传输在实际中还没用到。这里解释的主要是军事上的应用。

VHF 频段和 UHF 频段的主要应用当然还是在正常的和稳定的传播方式上。根据由准光传播特性所造成的较小的最大传输距离这一特点，人们可以在不同的空间范围内多次使用同一个频率。因此可以给宽带信道提供较高的传输能力。参考文献【Gen98】详细地描述了无线信道的建模。

无线广播占用了频率中的主要部分，参见表 5-2。飞行无线电台以及军事应用也在这个频段范围内。由于两个原因，蜂窝移动电话使用较高的 900MHz 的频率：其中一个原因是天线较小，这样设备可以是手持的；另一个是信号可以更好地覆盖城市的街道。

根据传播特性来区分 30MHz～1GHz 范围与高于 1GHz 范围并不如根据使用的技术来区别这两个范围更成功。在 30MHz～1GHz 这个范围内发送功率从几 W（移动电话）到几 kW（无线广播发射器）。天线使用的是圆形天线（棍状天线）和定向天线（多数情况是 Yagi 天线）。

5.5.6 1GHz 以上频段（微波）的传输

在这个频率范围发送功率只有几 W（特例：雷达和卫星）。主要应用是与无线广播相反的定向无线广播。即人们使用点到点的通信来代替一个平面覆盖，因此，多数情况下使用的都是定向天线（通常使用的是抛物线镜面）。

在定向通信中发送器和接收器之间具有可靠的连接。传播是无干扰的，因此也称为自由空间传播。在地面传输中必须注意，发送器与接收器的高度差要足够大，这样：

1）由于地球的曲率而造成的信号在中间落地不能超出传输路径。因此发送器和接收器的天线必须放在塔上或山上。

2）第一菲涅耳区（Fresnel Zone，参见下面）是无障碍的，因此要求额外的高

度差。

沿着传输路径高度差的值为

$$h=\frac{1}{2R}d(D-d) \tag{5-43}$$

式中，D 是传输距离；R 是值为 6378km 的地球半径；d 是长度变量，值在 $0\sim D$ 之间。

期望的是 h 与 $d=D/2$ 是对称的。最大的高度差出现在传输路径的中央（$d=D/2$ 处），值为

$$h_{\max}=\frac{D^2}{8R} \tag{5-44}$$

从 Konstanz（康斯坦茨）到边界的 Bodensee（博登湖）的长度（$D=46$km）产生的高度差已经为 40m！在几何对称的球体上必须将天线放在高度为 h 的塔上。通信距离 D 为可视距离 s 的两倍：

$$D=2s=\sqrt{2Rh} \tag{5-45}$$

30m 高的塔能提供的几何可视距离是 20km。

经过大气的衍射微波比光波达到的距离要远。人们称之为无线电水平线，并可根据式（5-45）通过将 R 放大 4/3 倍（对中欧是适用的）来计算。

由于地面的反射，多个波可以到达接收器。由此产生的干扰可能减弱总信号。因此所谓的第一菲涅耳区必须是无障碍的。这个区是椭圆形的，天线位于它的焦点上，如图 5-57 所示。经过反射的绕行路线到达椭圆形上时与直达路线刚好相差半个波长。如果在第一菲涅耳区内有障碍，那么直达信号与非直达信号（经过反射的）的叠加会产生破坏性的干扰。

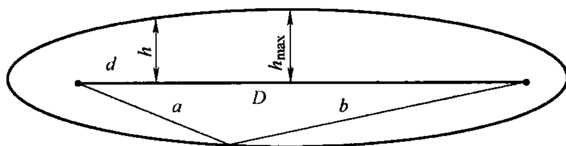


图 5-57　第一菲涅耳区 $(a+b)-D=\lambda/2$

图 5-57 中高度 h 的值为

$$h=\sqrt{\frac{d(D-d)\lambda}{D}}=\sqrt{\frac{d(D-d)c}{Df}} \tag{5-46}$$

高度的最大值仍然出现在中间位置，例如当通信距离为 50km 时，若频率为 5GHz，那么高度的最大值为 30m。

发送地和接收地地形之间的必要的高度差与两者之间的地区相比也与两个因素有关：地球的曲率和菲涅耳公式（5-43），以及衍射和式（5-46）：

$$h=\frac{1}{2\left(\frac{4}{3}R\right)}d(D-d)+\sqrt{\frac{d(D-d)c}{Df}} \tag{5-47}$$

其中两个加数都是关于 $d=D/2$ 对称的。因为这里的精度不是毫米，所以可以对式

(5-47) 进行一些变换，以得到经验计算形式。变量不是 D、d 和 f，而是采用 D、d/D 以及 $f/(1\text{GHz})$（归一化的频率）。除此之外 D 的单位用的是 km，其他所有的常数都综合在一起。这样会有

$$h = 0.06D^2\frac{d}{D}\left(1-\frac{d}{D}\right)+17.3\sqrt{\frac{D}{f}}\sqrt{\frac{d}{D}\left(1-\frac{d}{D}\right)} \tag{5-48}$$

这里 h 的单位是 m，D 的单位是 km，f 的单位是 GHz。

通信距离是 50km，频率是 5GHz 时，必要的高度差约为 70m。式（5-48）中各个值可以由每个地区的描述文件来控制。它由该地区的地形图（出现森林时给出的是地面的高度而不是这里感兴趣的树木顶端的高度！）或自动地根据数字化的地区模型来获得。必要时必须将发送地和接收地进行平移，或在中间放一个中继站。图 5-58 所示为对一个定向无线通信的几何/地形要求。

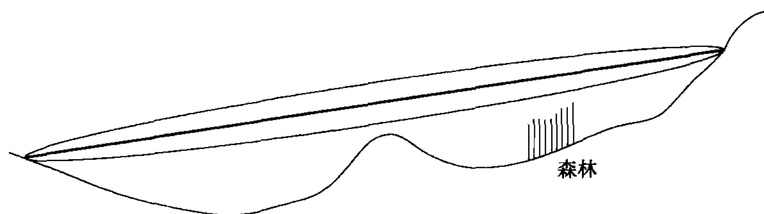


图 5-58 带有菲涅耳区的定向无线通信的区域概要文件

在自由空间传播时可以精确给出段落衰减。从根本上来说，它涉及的不是吸收而是稀释。一个波的功率密度可根据式（5-36）由 Poynting 矢量来给出。如果发送器发送功率为例如 1W，并将天线用表面积为 1m^2 的假想中的球形壳给罩住，这样在球形壳上的功率密度为 1W/m^2。将功率密度在球表面上做积分，那么结果是 1W，即闭合功率。将距离增大，这样假想的球的表面积就会增大，这样功率密度会下降，而闭合功率并不会下降。

无需推导：距离为 D 的自由空间上的衰减 A_F，在均质的（没有增益）发送天线和接收天线（即从天线连接到天线连接）上的值为

$$A_\text{F} = \left(\frac{4\pi D}{\lambda}\right) \tag{5-49}$$

对上式取对数即可得到自由空间的衰减系数 a_F：

$$a_\text{F} = 20\lg\frac{4\pi D}{\lambda} = 20\lg\frac{4\pi fD}{c}\text{dB} \tag{5-50}$$

式（5-50）可以换算成经验公式。为此，将频率 f 用 f_0，距离 D 用 D_0 来进行归一化：

$$\begin{aligned} a_\text{F} &= 20\lg\frac{40\pi fD}{c} = 20\lg\left(\frac{4\pi fD}{c}\cdot\frac{f_0}{f_0}\cdot\frac{D_0}{D_0}\right) = 20\lg\left(\frac{4\pi f_0D_0}{c}\cdot\frac{f}{f_0}\cdot\frac{D}{D_0}\right) \\ &= 20\lg\left(\frac{4\pi f_0D_0}{c}\right)+20\lg\left(\frac{f}{f_0}\right)+20\lg\left(\frac{D}{D_0}\right) \end{aligned}$$

现在例如设$f_0 = 1\text{GHz}$，$D_0 = 1\text{km}$，那么第一个加数的值为92.4dB，这样自由空间衰减的经验公式为

$$a_F[\text{dB}] = 92.4\text{dB} + 20\lg(f[\text{Ghz}]) + 20\lg(D[\text{km}]) \tag{5-51}$$

数值实例：在5GHz的频率上，距离为10km的衰减为126dB，距离为40km的衰减为138dB。

距离增加一倍，接收信号的功率会减少6dB（1/4）

这种关系与传播时是完全不同的，传播时是由吸收特性来刻画的，而这里用dB表示的衰减会随着距离的增加而增大。上面的数值例子说明，10km长的电缆（假想的！）上通信衰减为126dB，40km长电缆衰减是504dB。

在只有吸收物性的媒体上（铜线、光纤）信号随着距离的增加会呈指数形式减弱。而在自由空间传播中（稀释代替吸收）信号随距离的增加只是呈二次方形式的减弱。但根据式（5-51）在自由空间传播中有个较大的“初始衰减”。这意味着，在短距离传输中，有线传输需要较少的功率，而在长距离传输时显然要考虑无线传输。距离相等时衰减与所用的媒体有关，目前人们使用光纤在100km的距离内无需使用中间放大器。

当频率大于10GHz时式（5-51）给出的自由空间传播只在外层空间才适用，因为下雨所造成的吸收衰减是很明显的。60GHz时，空气中O_2所造成的吸收起很大的作用。

定向无线通信的功率投入如下：

$$P_{RX} = P_{TX} + G_{SA} + G_{EA} - a_F - a_K \tag{5-52}$$

式中　P_{RX}——用dBm或dBW表示的接收功率；

P_{TX}——与P_{RX}有相同单位的发送功率；

G_{SA}——用dB表示的发送天线增益；

G_{EA}——用dB表示的接收天线增益；

a_F——根据式（5-51）用dB表示的自由空间衰减；

a_K——用dB表示的天线馈电线、连接器等的衰减。

（通常还需增加3～6dB的“未分配系统衰减”）

脆弱的是发送器末端到天线以及天线到接收器之间的连接。这里人们使用的是同轴电缆（微波波段每米的衰减约为0.5dB，低频时较小）或更好的（但更贵、更不灵活）波导（每米的衰减为0.04dB）。在山上的定向无线传输设备由于天气的原因电子设备会与天线分离放置传输只有6m长的同轴电缆就能耗尽一半的发送功率！感谢小型化技术，目前微波电子器件可以直接放在抛物线镜面上，这样在馈线上只有中间频率信号（LNB及LNC，参见5.2.1节的结尾部分）或者甚至直接是基带信号。由于它们的频率很低，可以直接在衰减很小的同轴电缆上传输。

模拟传输时必要的接收信号功率P_{RX}最终由SR_A和接收器输出端希望的信噪比来确定，数字传输时由希望的误码率BER来确定：

模拟的数：

$$P_{RX} > P_N + SR_A - G_M \tag{5-53}$$

数字的数：
$$P_{RX} > P_N + SR_K - G_C \tag{5-54}$$
式中 P_{RX}——根据式（5-52），用 dBm 或 dBW 表示的接收功率；

P_N——根据式（5-42）的接收器的固有噪声，与 P_{RX} 有相同的单位；

SR_A——用 dB 表示的接收器输出端希望的信噪比；

G_M——用 dB 表示的调制增益；

SR_K——接收器输入端必要的信道信噪比，由调制类型和 BER 来确定，参见图 3-75；

G_C——用 dB 表示的编码增益。

对于定向无线通信来说，很明显应该选择尽可能低的频率，因为根据式（5-51）自由空间衰减随频率增加而增大，并且在低频上比较容易产生较大的功率。频率升高一倍，自由空间衰减会增加 6dB。但另一方面，根据图 5-45，天线的增益同样增加 6dB。当使用两个定向天线时，根据这一原则，值得使用尽可能高的频率。对远距离传输（几十 km）来说由于下雨造成的衰减频率的极限值为 10GHz。

定向无线传输设备工作时消耗的电能为 1W 或更少。通常天线的参数都超过要求的标准，这样做不是为了继续提高接收功率，而是为了使无线的辐射更好地捆绑在一起。这样同一个频率还可以在地理位置相邻的区域内继续使用。

通常情况下，会对式（5-52）进行一些变换。首先，发送天线的增益会与发送功率结合起来形成用 dBW 或 dBm 表示的 ERP（Effective Radiated Power，有效辐射功率）。其次，现代的定向无线传输设备会根据对方站点来调节自己的发送功率。因此，在式（5-52）中，给必要的接收功率 P_{RX} 增加一个如值为 10^{-8} 的 BER（这里被称为 P_{RXmin}）。这样可以计算所谓的系统值：
$$a_{Fmax} = \underbrace{P_{TXmax} + G_{SA}}_{ERP} + G_{EA} - P_{RXmin} - a_K - \text{预留衰减} \tag{5-55}$$
这个系统值只与设备有关，并可以根据式（5-51）给出区间衰减允许的最大值。衰减余量的选择是自己来定的，人们通常选择 10dB。要继续减小 BER，例如减小到 10^{-12}，最好使用信道编码而不是提高发送功率。

小结：

1）对微波链路的规划需要地形学的知识并使用式（5-48）进行控制，是否到达了必要的高度差。

2）因为衰减而引起的最大传输距离根据式（5-55）和式（5-51）来计算。如果衰减太大，就要使用例如较大的天线。

3）应该计算大概的接收功率（以及可调节运行下的大概的公式功率），这样在安装天线时不会影响副辐射瓣的最大值。除此之外，这些值还可以作为参考值，允许对这些设备进行运行时的监测。

在规划整个定向无线通信网络时，对每条链路都要做如上所述的预算。此时涉及的是，除了地形学，不能超过规定的衰减的最大值。

对于一个网络，还必须保证任何两条链路之间的衰减要超过规定的最小值，这样能够消除相互之间的干扰。衰减与下面因素有关：地形、天线辐射图、极化衰减为20～30dB

时的不同形式的极化、不同的工作频率和接收滤波器的倾斜度、使用的反射功率。

手工网络规划只能是大致地执行。计算机辅助方法能更好地利用频谱。能够自动调节发送功率的站点允许构成更密集的网络，因为它减少了同信道干扰。

定向无线通信部分可以使用622MBit/s 的数据传输速率，是使用同轴电缆或光波导的有线传输的替代方式。与有线传输相比的一个最大好处就是较短的链路安装时间以及信号在传输路径上的相对不易破坏性。缺点是设备需要安装在较好的位置上，不符合环境保护者的要求。长距离通信（100km）在 4 ~ 10GHz 的频率范围内进行，而短距离通信（几 km）可在 10 ~ 60GHz 的范围内进行。

发展现代定向无线通信系统的目的是：

（1）改进传输质量（提高无差错的比特率）

多径传输和频率选择性衰减会导致线性失真，对传输产生负面影响。它会引起码间干扰（ISI）。克服 ISI 的办法是利用空间及频率的多样性以及使用可适配的均衡器。

（2）提高经济性（提高单位花费的比特率）

人们越来越多地尝试使用高集成度的电路在基带内进行信号处理，因此使用如图5-28或图 5-29 所示的数字接收器。继续产生影响的是网络的融合和网络管理功能的提供（第 6 章）。

（3）提高频谱有效性（提高每 Hz 的比特率）

定向无线信道由 ITU-R 标准化，带宽是 28 ~ 40MHz。为了能有较高的数据传输速率，多级调制是必要的。目前已经实现了使用栅格编码的 64QAM 到 256QAM 系统。

共信道操作（cochannel operation）是提高频谱有效性的另一种方法。这里两个定向无线信号使用相同的频率来传输，但这两个频率使用相互正交的极化方法，如图5-59所示。

百分之百的信道分离在共信道操作中是不可能实现的。因为传播效果（下雨）和天线的公差会导致一定程度的极化因而产生一定的串扰。使用交叉极化均衡器可以成功地消除这些串扰，如图 5-60 所示。这样串扰衰减可以达到 0dB，即有用信号和干扰信号同样强！交叉极化均衡器的工作原理同图 2-20 所示的回波抵消器相同。

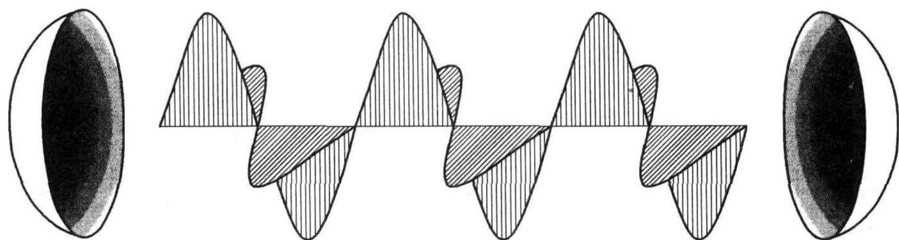


图 5-59　两个定向无线信号使用同一个信道

对微波设施小心是值得的。在这个频段内辐射虽然不会电离，因此不会引起癌症或对基因有危险，但它会使眼睛变得模糊（蛋白质凝固）。因此人们必须对大功率密度的微波多加小心。安装好的频率为 5GHz，功率为 1W 的发射器是没有危险的。这种设

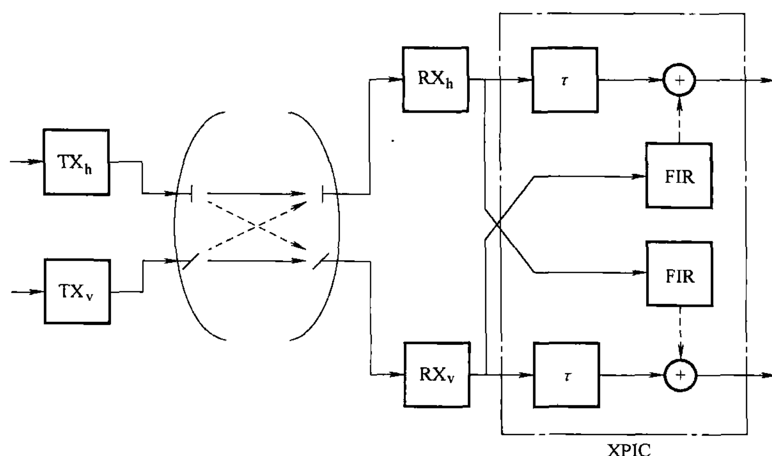


图 5-60 交叉极化均衡

施使用直径约为 1m 的抛物线形天线。如果将天线的电缆取下并仔细观察电缆的端点，这时的功率密度会比正常操作时大很多。

最后通过数值实例来明确一下微波链路各参数的确定过程。假设我们希望实现一个速率为 2MBit/s，距离为 50km，BER 为 10^{-8} 的定向无线传输，地形上是允许这个通信的，因为两端的高度差可以保证足够大。

首先要确定频率。为此要考虑国际和国家的频率分配计划。这个例子选择频率为 5GHz。

根据式（5-51）计算得到的自由空间衰减为 140.4dB。这个衰减可以用天线增益、发送功率、调制增益和编码增益来补偿。因此有多种可能的系统解决方案。对这个多样性进行一下限制，这里不使用信道编码，并首选抛物线形天线。天线的直径应该为 1m，这样在一定程度上还允许手持。根据图 5-45 或式（5-38），每个天线的天线增益为 32dB。

现在来选择调制类型。这与可用的带宽有关。假设对带宽没有限制，因此选择 OQPSK。根据图 3-75，这需要一个 12dB 的 E_{bit}/N_0。根据式（3-60）和式（2-26），计算出卷积因子为 0.35，因此 SR_{K} 为

$$\frac{P_{\mathrm{S}}}{P_{\mathrm{N}}}=\frac{E_{\mathrm{Bit}}}{N_0}\cdot\frac{R}{B_{\mathrm{ü}}}=\frac{E_{\mathrm{Bit}}}{N_0}\cdot\frac{R}{\frac{R}{2}(1+r)}=\frac{E_{\mathrm{Bit}}}{N_0}\cdot\frac{2}{1+r}=\frac{E_{\mathrm{Bit}}}{N_0}\times 1.48 \tag{5-56}$$

系数 1.48 意味着 1.7dB。因此 SR_{K} 的值至少为 12 + 1.7 = 13.7dB（比较 3.4.5 节末尾的数值实例）。

现在根据式（5-42）来计算接收器的噪声。假设接收器的噪声系数为 10dB，天线的噪声系数为 1dB，这样 P_{N} 为

$$P_{\mathrm{N}}=-174+10\lg(B)+10+1=-163+10\lg\left[\frac{R(1+r)}{2}\right]=-101.7\mathrm{dBm} \tag{5-57}$$

现在根据式（5-54）来得到必需的接收电平：

$$P_{RX} > P_N + SR_K = -101.7 + 13.7 = -88\text{dBm} \quad (5\text{-}58)$$

最后根据式（5-52）来计算必要的发送功率。这里假设微波电子元器件直接安装在天线的镜面上因此没有电缆上的损失：

$$P_{TX} = a_F + P_{RX} - G_{SA} - G_{EA} = 140.4 - 88 - 32 - 32 = -11.6\text{dBm} \triangleq 70\mu\text{W} \quad (5\text{-}59)$$

这里对天线偏差、雷雨天气、微波部件的接口等所造成的损失没做一点预留。根据连接的重要性应该考虑做不同的预留，使用发送功率调节器可以在必要时激活这些预留。功率为 10mW 的发射器在这里可以提供超过 20dB 的预留，在 5GHz 频率上可以毫无问题地进行安装，整个系统也会工作。甚至可以考虑减小天线的尺寸，继续提高发送功率。

说明：在这个数值实例中 dB 的数值给出了小数位。从这个角度上来看，较大的预留过于精确。尽管如此还是应该这样计算，这样可以了解预留，不会将它丢掉，并且下意识地将它放在许多舍入中。

如果还另外加入信道编码，那么可以这样计算，它可以改善 BER，例如提高了 100 倍。这样 SR_K 的最小值会下降，为此码速率下降时（即提高了冗余）带宽会升高。根据这个规则，结果会是发送功率的最小值会减小，否则的话采用复杂的信道编码就不值得了。

5.5.7　卫星技术

卫星指的是人造的地球卫星，它为各种各样的科学、技术、经济和军事上的目的而服务：

1）通信卫星：无线广播卫星（电视和广播的直接接收）和通信卫星（商用网的远程传输，不能直接到达最终用户）。

2）地球和宇宙空间的探索卫星：如气象卫星、资源探索、气象观测、间谍和侦查卫星、用于天文观察的哈勃望远镜等。

3）特殊用途的卫星：如导航。

本节的主题是通信卫星。对通信技术人员来说，这些卫星首先是具有特殊性质的正常的中继站点，它们的特点就是被放在宇宙空间中。由于它们的特殊高度，卫星能够俯瞰地球表面的绝大部分地区，能够保证一定质量和可靠性的宽带、远距离通信。正是这个特性使卫星通信比短波通信更强大：后者是窄带的，并且由于电离层的时变性通信是不可靠的。卫星通信的缺点是技术上很复杂，必须依靠第三者，即拥有太空技术的国家。

卫星包含多达 100 个的所谓的异频收发器，它们的结构如图 5-61 所示。一个异频收发器使用混频器将输入端的频率范围转换成输出端的频率范围。混频器之后又充分利用反向频率点以补偿多普勒效应，参见 3.1.3 的总结评论。

由异频收发器转换的频带的带宽为 36～50MHz。这个物理信道可以根据下面的方法被分成多个逻辑信道

1）频分复用（FDM）；

2）时分复用（TDM）；

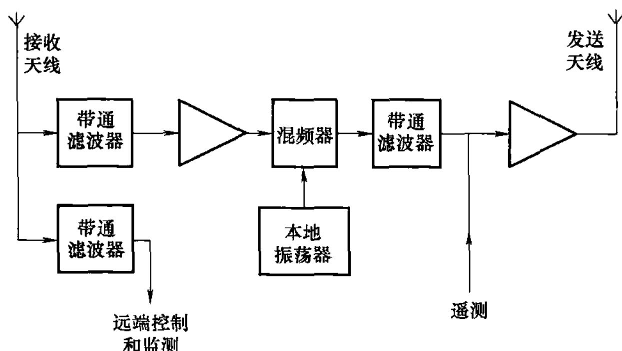


图 5-61 卫星异频收发器的组成原理

3）TDM 和 FDM 相结合的方法；

4）码分复用（CDM，扩频）。

由于电离层的衍射从地球到卫星的链路（uplink）和从卫星到地球的链路（下行）的频率在 20～30MHz 上。在实践中使用的频率在 1～30GHz 这个范围内，有些卫星使用的是从 100MHz 开始的频率范围。超过 10GHz 由大气所造成的衰减已经起很大的作用了，不论是上行链路还是下行链路都要根据 5.5.6 节所介绍的方法计算各种参数。相反陆地上的定向传输保证非涅耳区的畅通并不是很难的。根据式（5-52）链路的支出必须要能填补第二个段落。这个支出可以很容易地采取图像化的方式用电平或是衰减来描述，图 5-62 给出了一个实例。它的原理同图 1-20 给出的是完全一致的。这里给出了两个卫星之间的连接（interlinks），当然可以采用相同的计算方法。

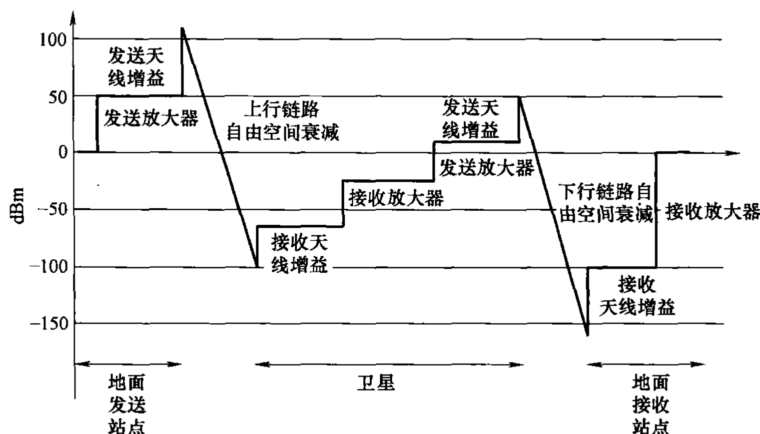


图 5-62 卫星连接的电平图：地面天线增益 60dB，卫星天线增益 40dB，自由空间衰减 210dB（频率为 15GHz，距离为 40000km）

图 5-61 给出的简单的卫星原理包含几个巧妙之处：

1）传输距离和由此相关的自由空间衰减与陆地的定向传输相比是很大的。产生这

个结果的前提条件是强发送功率放大器、小噪声接收放大器以及大增益天线。这个天线要求的张角很窄，这样天线才能正好对准在高轨道中飞行的卫星。对于在较低的轨道上的卫星（Low Earth Orbit，LEO），天线的聚焦是没有问题的，因此此时不必有较大的增益。

2）卫星的位置：卫星离地球越远，卫星在地球上的可视范围就越大，因而能够连接的地球上两点的距离就越远，地球上的某一点能见到卫星并因此能够使用卫星的时间就越长，信号也就越弱。这里也需要在各方面因素之间寻求一个最优折中。后面还要对卫星轨道的选择进行特殊说明。

3）卫星中的电子设备：卫星使用太阳能电池来供电，它们位于卫星表面或镶嵌板上，只有到太空中才展开。由于机械上的复杂性，这些镶嵌板不能做得太大，因此卫星会有供电的问题。各种电路必须采用高有效系数的方式进行实现。

太空中缺少在正常情况下能将电子电路产生的热量带走的空气。因此卫星上的电子电路对于局部过热是很危险的。使用特殊的导热管能将功率放大器产生的热量导向卫星的外表面。遗憾的是具有较好的热传导性的固体它的导电性能同时也很好。

4）卫星的机械特性：面对太阳一侧和背对太阳一侧的巨大温差给卫星的表面造成很大的压力。任何可移动的部件都要特别注意，因为润滑在真空中实现起来并不是很容易的。因此，例如由于天线或太阳能板不能伸展而造成的卫星丢失并不是很少见的。

5）卫星的稳定性：卫星作为一个物体围绕地球运动，即只有卫星的质量重心是定义的。不采取特殊的稳定措施，卫星就会在太空中旋转或摇晃，这使得对于面向地球的定向天线和面向太阳的太阳能板的最优调准是不可能的。为了保持卫星的稳定，使用了不同的措施：老的卫星使用旋转稳定的办法（卫星围绕它自己的轴每分钟旋转10～30次），这只允许定向天线有一个轴，并且对卫星上的光学仪器是不适合的。在低轨卫星（LEO）中使用根据地球的重力场或磁场来调节位置的方法，而对远轨卫星通常利用恒星来调节。其他卫星使用自带的磁性的无接触的陀螺仪。在通信卫星的大家族中，例如有的陀螺仪质量为3.5kg，每分钟自旋3000转。

卫星上携带有小型的助推器，用于对位置和轨道进行调整，它只是在必要时才被短时间激活。目前这个助推器的燃料存储量足够整个卫星使用的时间。

图5-61所示的异频收发器只给出了频率转换和放大功能。对数字信号来说还需要一个再生器。图2-18给出了一个基带转发器的结构，在BP传输时还必须添加一个解调器和调制器。对较高的要求只需要上行链路或下行链路分别使用信道编码即可。对没有实时要求的（例如E-mail）数字传输中，在卫星上还可以扩展一个存储器。在这种情况下，卫星就是一个飞行的邮箱，如图5-63所示。

一个很有意思的应用是使用低轨卫星，因为距离短所以使用较低的复杂性它就能实现。卫星飞过一次所持续的时间只有10～15min，可见范围很小。一个大文件，例如从欧洲发往非洲，被分成多个小块，当卫星飞过一次或多次时将这些小块发往卫星上的存储器中。当飞过非洲上空时，存储器中的内容被转送到陆地上的一个盒子中，并从那里被继续转发。

现在我们来说一下卫星的轨道（Orbit）。理想情况下（地球是质量均匀分布的球

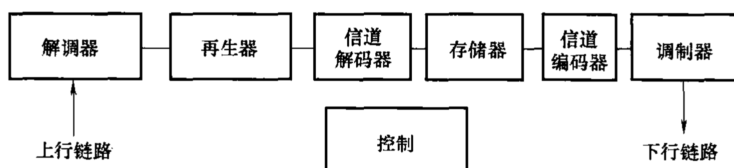


图 5-63 卫星邮箱

体，太空中没有太空垃圾，不受太阳和月亮的引力场的影响等）卫星轨道可以完全通过开普勒定律来计算。根据这个定律，卫星围绕地球在一个椭圆形轨道上旋转，地球位于这个椭圆形轨道的一个焦点上，如图 5-64 所示。旋转轨道是一个特殊的椭圆形轨道，它的两个焦点相互分离。地球的引力与一个离心力相互抵消，即卫星离地球越远，它相对于地球移动的越慢。因此在椭圆形轨道上角速度是变化的。

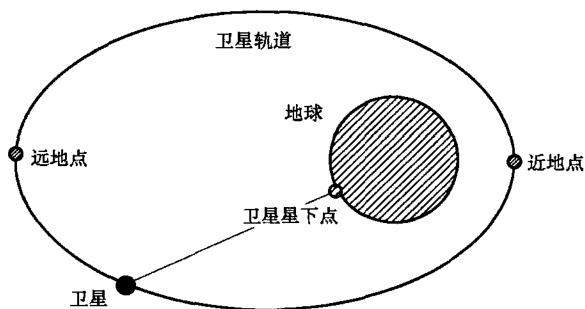


图 5-64 卫星的轨道

椭圆形由离地球最近的点（近地点）、离地球最远的点（远地点）、偏心距以及椭圆平面和赤道平面的夹角来确定。

轨道平面（椭圆平面）在开始时就被确定，只有特殊的能量消耗才能改变它。这是因为速度矢量在轨道平面的分量很大。如果人们还是希望改变轨道平面，例如将卫星射向深层宇宙空间，那么可以利用其他星球的引力场。

但在轨道平面内部，椭圆形相对容易被改变。为此需要使用卫星上的一个发动机在特定的时间上给卫星加速。如果在远地点加速，椭圆形会变得很大，旧轨道的远地点就会成为新轨道的近地点，如图 5-65 所示。为了节省发射火箭的推进能量，会充分利用地球的旋转速度。正是这个原因，几乎所有的宇宙空间站都在赤道附近，并且起始方向向东，即地球旋转的方向。

椭圆形旋转轨道的好处是，当卫星在远地点附近时，在地球表面上的可视时间长并且范围大。如果将远地点放在所关心的地区，那么卫星在较长的时间内都是可用的。

由于地球本身是旋转的，卫星每次转过时看到都是地球上不同的区域。如果将椭圆形旋转一周的时间选择为 24h，那么卫星每天都覆盖相同的区域，这是一个地球同步轨道。遗憾的是由于地壳的扭曲，轨道平面产生运动，这样上面所描述的轨道的远地点从地球上来看是变化的。人们可以证明，当轨道平面的倾斜角是 63.4°时，轨道平面

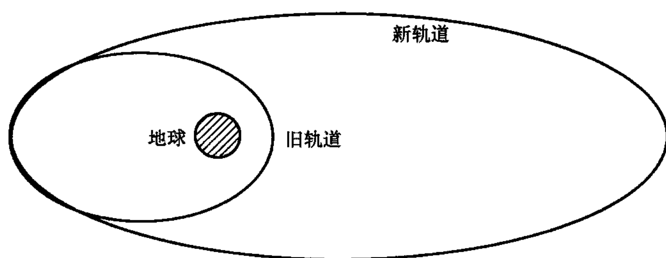


图 5-65　卫星轨道高度的改变

的运动从地球上看是不可见的，这样轨道真正变成了与地球同步。这个轨道首先是由俄罗斯的 Molniya 系列卫星使用的，因此具有 63.4°倾斜角，24h 旋转周期的椭圆形轨道也被称为 Molniya 轨道。这样俄罗斯人可以为它更北部的、使用一般的对地静止卫星很难到达的西伯利亚地区提供服务。

地球同步轨道（旋转时间是 24h）的一种特殊情况是对地静止轨道，即从地球上看卫星始终停留在同一位置。这样唯一一种可能性是圆形轨道距离地球 36000km，倾斜角为 0°。所有的地球同步卫星都位于赤道的上空，覆盖地球上大约 1/3 的地区，无需天线跟踪控制。因此这些轨道特别适合无线广播卫星。对欧洲提供服务的对地静止卫星距离中欧的距离约为 40000km，它的位置由子卫星点的经度给出，例如是西经 19°。

对地静止轨道很受欢迎，因此也很拥挤。相应地在那里也积累了许多宇宙垃圾，因此也会有这样的可能性，未来的卫星通信会特别的贵！

对地静止轨道也不是精确地对地静止的，事实上一次可注意到的卫星运动会产生大约 600km 的距离。为此卫星之间需要保持一个安全距离。这些运动从地球上看会产生 1°的张角范围。用于比较：满月产生的角度是 0.5°。根据图 5-45，直径为 2.5m 的抛物线镜面在 8GHz 时张角约为 1°。经过很强复用的天线也必须要跟踪对地静止的卫星。相反，人们在无线广播卫星上安装较强的发送功率放大器，这样在接收端需要较小的天线就可以。这些小天线可以固定安装，不引人注意，当然也比大天线更便宜。

对地静止轨道的最大优点是卫星的位置是固定的。缺点是距离很远，需要复杂的技术设备。对无线广播卫星来说，复杂工作都集中在卫星上，而对于通信卫星来说，复杂工作会集中在地面站点上。无线广播卫星边长约为 2m，质量可达数吨。

另外一种极端情况是发射较小的低轨卫星（LEO，即高度最大为 1500km，运行一周的时间约为 2h），不久的将来还会听说关于这种卫星的更多的内容。通过这些卫星，手持无线设备就能使用卫星业务。

LEO 在很低的两极的允许轨道上（倾斜角为 90°），具有较好的观察解析度和敏感性，利用 footprint（在地球上的覆盖区域）可以一步一步地覆盖整个地球。这些轨道适合于地球探测卫星。

在卫星上安装天线需要精确的几何数据。这可以由所谓的卫星轨道的开普勒元素以及大量的球面三角学来获得。因为卫星轨道是不断变化的（地球引力场的不均匀性，

太空垃圾所造成的减速，太阳和月亮重力所造成的影响)，要对卫星轨道进行周期性的测量，并要对开普勒元素进行实时更新。天线调准的计算工作可用舒适和便宜的计算机程序来完成。

必要时需要对卫星的轨道进行校正，为此需要使用已经提到过的很小的控制助推器和图5-61所解释的遥控。轨道校正由地球上的控制中心来完成的，而卫星的位置校正由卫星自己来完成。

每颗卫星都会偶尔地处于地球的阴影中。例如对对地静止卫星来说这通常发生在春季和秋季开始时。从卫星上来看这就是一次日食，是一个很微妙的阶段。在这个阶段中，能量的补充没有了，因此必须使用电池来渡过这个阶段。在这个阶段中，卫星上的发射器被关掉了，以便能延长蓄电池的使用时间。另一种特殊情况是从地球上看卫星引起了一次“日食”。从视觉上这当然不会出现，但对下行链路的接收天线来说，除了卫星信号之外，还有额外的太阳的噪声，会引起信噪比的剧烈下降。

作为调制类型，在模拟传输中FM和窄带FM用的很多，这是因为卫星上的功率放大器的有效系数的缘故。对于数字传输，OQPSK方法（见图3-69）很受欢迎，这是出于对有效系数、带宽效率和抗干扰性的考虑。卫星的发送功率根据类型不同值从几瓦特到约250W，后者是给终端用户直接接收的无线广播卫星上的值。无线广播卫星的功率需求约为2～3kW。为了充分利用带宽，使用了如图5-59和图5-60所示的交叉极化的共信道系统。多个MPEG-2编码、复用成33MBit/s的视频流可同时使用相同的卫星信道来传输，并且能用直径只有60cm的抛物线形天线来接收（DVB-S，参见5.1.2节）。

通信卫星被用来进行数据传输和演播室之间的视频通信，例如“Eurovision广播”。以前经常是多达15000个电话通信同时通过一个对地静止的卫星来传输，例如经过一系列的国际通信卫星。现在它们已经被跨海洋的光波导所代替，因为在超过2×40000km的距离上大约270ms的传输时延对话音来说是很不舒服的。在有线传输中每1000km只有5ms的时延。通过卫星的数据传输速率可达155MBit/s，将通信卫星连接到SDH网络上是可行的（SDH为同步数字序列，是一种复用的方法，在6.3.2节中会做介绍)。

对卫星来说很危险的是发射阶段（振动，火箭的离开）对运行轨道的影响（推动燃料的使用和由此产生的运行时间的减少）以及天线和太阳能板的展开。如果一切都很幸运的话，卫星可以使用10年，之后，用于校正轨道的助推燃料会用尽。小卫星人们会让它停止并让它自己在大气中燃烧掉，大的卫星人们会让它向太阳的方向飞去。卫星的可靠性（可用性）超过99%！

现代的卫星有多个异频收发器和多个天线。它们能用所谓的点波束向直径为几百千米的小范围内辐射。目前在太空中放置的异频收发器的数量估计超过2000个，预测说未来几年的它的增长率为50%。

以前卫星都是独立奋斗的，现在出现了卫星网络。它在以下几种情况时提供服务：

1）使用统一的系统提供世界范围内的服务

2）陆地通信很难包括的范围，如印度和中国（还比较差的基础设施，但有较大的

通信需求)、印度尼西亚（有 13677 个岛屿)、澳大利亚（很大，没有包括的沙漠地区)。

Inmarsat 组织是不同国家电话公司的一个联盟，它维护一个卫星系统，它的公文包大小的 C 版本关于使用对地静止卫星的地面站数据通信的标准被通过。除了南北极的冰帽，地球上的所有地区都被覆盖。使用与陆地电话网的转接器，Inmarsat-M 提供 4.8kBit/s 的话音通信（LPC 压缩）或 2.4kBit/s 的数据和传真通信。

使用地面站，大量的、用户友好的 GSM 手机还不能直接与 40000km 之外的对地静止卫星进行通信。取而代之的是使用低轨卫星（LEO)。它们从地面上来看移动的是非常快的，因此在一个通话期间，会从一颗卫星上切换到另一颗卫星上。原则上它涉及的是与 GSM 技术相同的蜂窝网络（参见 6.5.2 节)，但此时蜂窝是移动的而不是用户移动。每颗卫星用点波束形成了多个无线蜂窝。作为例子，我们简单解释几个系统，这里涉及到的如 TDMA 和 CDMA 等接入方法将在 6.2.1 节中进行解释：

1）Iridum：从 Motorola 引入的系统，由 66 个卫星组成，每颗卫星重约 700kg，高度为 780km。因此每颗卫星最长只有 13min 的时间从地球上是可见的。最初设计的是 77 颗卫星，系统的名字就来源于原子序号为 77 的化学元素。接入技术使用的是 TDMA，上行链路使用的是 19.5GHz，下行链路使用的是 29.2GHz。1998 年初正式开始使用，投资估计是 50 亿美元。商业上，Iridum 系统是失败的。Iridum 系统是为覆盖 GSM 系统并连接孤立的地区而设想的，但却未想到与分布很广的陆地网络产生了竞争（可能的用户还在那里，但从他们那拿不到钱!)。除此之外，数据传输速率是为话音通信而设计的，对其他的业务来说太小。人们都想放弃卫星，将它们燃烧掉，但美国国防部可以买这个系统。

2）Inmarsat-P：由 Inmarsat 运营的 12 颗卫星组成，在地球上空 10000km 处，支持的通信同 Inmarsat-M 相同，但使用手机作为地面站。从 GSM 网络内使用多模手机也能使用该网络。接入技术是 TDMA。

3）Globalstar：高度为 1400km 的 48 颗卫星，在 8 个平面内运动，每个卫星的作用都是一个纯转发器，它们使费用低廉的通信成为可能。另外 8 颗卫星已在轨道上备用，还有 8 颗在地面上备用。接入是使用 CDMA 来实现的。上行链路频率为 1.6GHz，下行为 2.5GHz。支持该项目的公司有：France Telecom、Alcatel、Hyundai 等。花费是 27 亿美元。

4）Teledesic：这个项目能够出现并引起轰动，都是因为有 Bill Gates 做它的教父。840（!）颗卫星应该在地球 640km 处运行，后来这个项目变成 288 颗卫星在 1600km 处运行。数据通信应该使用 ATM（参见 6.3.5 节)。估计投资约 120 亿美元，计划完成时间为 2003 年。

使用 LEO 的卫星相互之间一定要能通信。为此在相同的轨道上的卫星使用 ISL（Inter-Satellites Links）通信，而位于不同轨道上的卫星使用 IOL（Inter-Orbit Links）进行通信。IOL 还必须克服多普勒效应。因为在通信路径上没有大气，因此频率的选择是向上自由的。在 Iridium 系统中 Interlink 的工作频率是 23.3GHz，而新的项目正在研究使用光通信，因为那里的复用是最强的。一个实验系统使用下面的数据工作：传输距离为

45000km，激光功率为60mW，波长为850nm。光瓣只有600微度宽，在45000km之外的目的地，被照亮的部分的直径只有150m。而在辐射的这段时间内，卫星只移动了2km，人们需要一个精密的系统来控制辐射的方向。这需要事先的计算，粗校准和精校准。多普勒变化是10GHz，敏感性是当数据速率为50MBit/s时，每Bit 73Photonen。

5.5.8 小结

图5-66总结了不同无线传输的应用范围。这里所给出的数据速率和距离只是参考值。

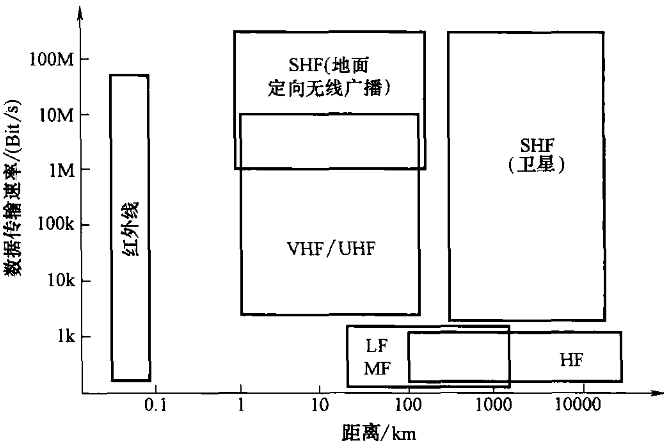


图5-66 无线传输的使用范围（频率范围的命名参见第80页中的表1-10）

5.6 有线传输媒介

有线传输是最老的电信传输方式，但它一直都很有效。最开始使用的是双绞线，后来有了同轴电缆，再后来又有了光波导（LWL）。各种线路的缺点都是线路的铺设和维护开销巨大。在一根电缆中可以有多对线（多个同轴管以及LWL），这样可以提高电缆的容量。这被称为空间复用。

与在4.3.6节中提到的空间多样性有些不同：两条线路经过不同的路径传输同样的信息。多样性是为安全性来服务的，而复用（即多重使用）能降低费用。

5.6.1 双绞线

双绞线（英文：twisted pair）意味者将两条铜线相互缠绕在一起，以便通过磁场来减少散杂干扰。电场之间的相互干扰有助于形成屏蔽。因此双绞线被分成STP（Shielded Twisted Pair，每对线单独进行屏蔽，只被IBM使用）和被广泛使用的UTP（Unshielded Twisted Pair，整个电缆一起做屏蔽）。两种类型的双绞线都使用空间复用，至少是4对线。

双绞线被用在 LAN（Local Area Network）和电话网中连接终端设备与第一个交换机（接入交换机，图 1-49）的用户接入线中。这个所谓的用户范围（local loop，subscriber loop）形成了很细的分枝。总的来说由于单个用户在空间上的分离，复用是不可能的。因此在本地环路上必须使用尽可能便宜的传输媒介，并且当数据传输速率较低时能做到在数千米的距离上不使用中间放大器而完成传输的任务。

在电信网向数字化过渡时，用户范围的传输线无需进行改动，这意味着巨大的经济效益。世界范围的电话网中约 2/3 的花费都在用户范围内。由于距离较短，144kBit/s 的 ISDN 使用双绞线可以很好地完成任务。但将来具有较高数据需求的应用（如多媒体）还需要很高的数据传输速率。HDSL 就提供了一种解决方案（High Speed Digital Subscriber Line，高速数字用户线，参见 5.6.4 节）。

与同轴电缆和 LWL 相比，双绞线是最便宜、最容易管理的媒介。但为此数据传输速率和传输范围都受到限制。

即使是最简单的 UTP 电缆，质量也是不同的。因此人们将它按照表 5-8 分成了不同的类别。

UTP 电缆的波阻抗是 100Ω，而 STP 电缆是 150Ω。对话音传输来说（local loop）使用 UTP 电缆的第 1 类和第 2 类，而在数据传输时（LAN）使用较好的电缆。

表 5-8　UTP 电缆的技术数据

根据 ISO 的分类	UTP 类型	最大频率 /MHz	最大衰减/dB (频率单位为 MHz)				最小近端串扰衰减(NEXT) /dB(频率单位为 MHz)			
			0.77	16	20	100	0.77	16	20	100
A	1,2	0.1	—	—	—	—	—	—	—	—
B	3	16	2.23	13.1	—	—	43	23	—	—
C	4	20	1.87	8.85	10.2	—	58	38	36	—
D	5	100	1.80	8.20	9.18	22.0	64	44	42	32

给一个大公司布线是很昂贵的。这要通过所谓的通用通信布线（UKV）来完成，UKV 也是被国际标准化的（ISO/IEC 11801 和 EN 50173）。UKV 有两点是特别重要的：

1）无源部分（电缆）与有源单元（集线器、转发器、交换机等）分离。电缆比设备能多生存好几代。

2）有共同起源的群和覆盖大面积的布线。由于它的灵活性，为整个网络生存期而计算的企业花费会降低，因为这样做之后，对组织机构的重组和新的通信需求的适应性会变得好些。

传统的 UKV 是星形结构，并包含 3 个层次：区域、大楼和楼层。图 5-67 描述了它的主要原理。对于一个楼层内的布线，目前只使用类型 5。其他层次根据数据传输速率的要求使用 UTP 电缆或 LWL 来实现。

目前正在讨论引入 UTP 类型 6。但问题是工作位置上是否需要 155MBit/s 的速率（相当于 ATM 技术中的能够提供多媒体应用的宽带信道）。比较明智的可能是，提高给工作位置上提供的信息的质量而不是数量，使用较好的信源编码。

5.6.2 同轴电缆

同轴电缆被用在下面的应用中：

1）电话网中的频率复用传输：带宽为60MHz时使用SSB能提供10800个话音信道，中间放大器的距离仅为1.5km。目前人们使用的是LWL中的TDM，而不是同轴电缆中的FDM。作为中间阶段，在同轴电缆上使用TDM，数据传输速率可达140MBit/s。

2）有线电视（CATV）：60个电视频道和几十个广播节目可用600MHz的带宽送到家里。但距离还有点短，因此目前CATV网络是使用LWL大致布署到住宅区域内，然后再使用同轴电缆精细地布署到每个房间内。

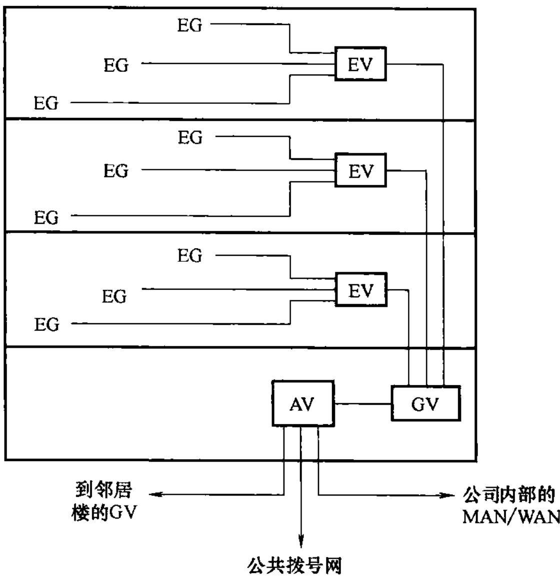


图5-67 通用通信布线原理（UKV）
EG—终端设备 EV—楼层分配器 GV—楼内分配器
AV—区域分配器 MAN—城域网
WAN—广域网

3）数据网（LAN）：考虑使用UTP，因为距离为1km时它能传输的速率为100MBit/s，距离短时，速率值甚至可以更高。

同轴电缆有较高的带宽，有一定的抗干扰性。抗干扰性是屏蔽的作用，由于集肤效应，它在60kHz以上才起作用。

集肤效应的另一个结果是所有铜线电缆随着频率的升高衰减都会加大，参见表5-9。因此当数据传输速率超过100MBit/s时，同轴电缆被光波导所代替。

根据构造方式的不同，同轴电缆的波阻抗是75Ω或50Ω（数据传输，专业无线技术时的天线馈线）。

表5-9 几种同轴电缆的衰减值

类型	Z_W/Ω	外直径/mm	衰减/(dB/100m)				
			10MHz	30MHz	100MHz	200MHz	500MHz
RG-58	50	5.8	5	9	17	24	39
RG-213	50	10.3	2	3.7	7	10.2	17
RG-220	50	28	0.6	1.1	2.3	3.8	7

从1.1.8节中我们已经知道，信息、带宽和能量是相互联系在一起的（香农立方体）。光波导由于带宽很大，可以传输很多信息。但同轴电缆却适合于传输很多能量。这也解释了前面遗留下来的使用轴电缆的一个应用的原因：作为发送器或接收器与天线之间的连接。

由于随着频率的升高衰减会增大，因此同样的任务在微波领域（大于 3GHz）会用波导（空心波导管）来完成。但这是很贵的，并且不够灵活，实际运行起来并不是很容易。例如在空心波导管内部不允许有水蒸气，否则的话衰减会有很大的增长。因此人们通常将波导放在氮压力较大的地方。

5.6.3 光波导

1. 引言

随着应用技术的发展，光波导（LWL）以及光电转换器（激光二极管和光敏二极管）逐渐成熟，并在正确的时刻内出现，它是电信技术的一个幸运物。从 1970 年开始就预测到的数据传输速率需求，因为衰减所以不能使用同轴电缆来实现，也因为衰减以及较小的带宽所以不能使用双绞线在可忍受的复杂性内（中间放大器的数量和距离）来实现。在危险的时刻人们不断寻找救命稻草，研究使用超导体来进行长距离的信息传输。

LWL 基于多种不同技术的组合，物理、材料科学和电路技术相互补充。幸运的是每个领域都在正确的时间作出了应有的贡献。结果就是 LWL 在宽带远距离传输领域代替了同轴电缆。

光传输的优良特性是：

1）速率可达 10GBit/s（受光电转换器的限制，光纤本身的传输可以超过 50TBit/s）。

2）低于 0.3dB/km 的小衰减，与表 5-9 中的衰减值相比有巨大的区别！

3）抗电磁干扰的能力。

4）很好的防窃听能力。

5）与同轴电缆相比，体积小，重量轻。

图 5-68 所示为目前所有光传输系统的原理：



图 5-68 光传输系统的原理图

与图 1-24 做比较，可以将图 5-68 中的转换器简单地理解成调制器和解调器。实际上目前用户无需光电子知识也能实现 LWL 传输。系统可以简单地根据传输结果来判断，而不是根据它的作用方式，即根据信息信号的类型（模拟或数字），1.1.7 节所解释的评价准则都是适用的。

尽管如此，电信技术人员了解一些物理方面的知识还是很明智的。因此本部分可以看成是一种关于 LWL 传输的“management summary”。

2. 光纤

光传输基于光从光密介质到光疏介质传播时会发生全反射的原理。因此 LWL 可以看成是一组按照同轴的方式排放的具有不同折射指数的玻璃，如图 5-69 所示。LWL 的基本材料，即硅（石英玻璃，SiO_2），不夸张地说像海洋中的沙子一样多。

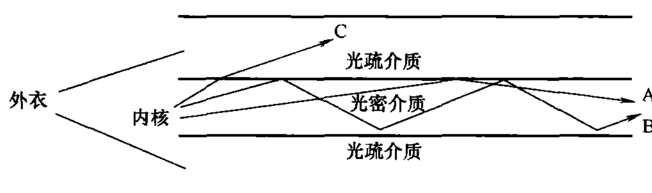


图 5-69 通过 LWL 的纵断面，由内核（core）和覆盖层（cladding）组成，带有 3 个光辐射 A、B 和 C（没有画出在机械上进行保护的外衣层（coating））

光辐射必须足够平地进入到光纤中，这样才能产生全反射。图 5-69 中的辐射 A 和 B 就满足这个条件，但 C 不满足。极限角度的正弦值被称为光纤的数值孔径，它与几何尺寸（内核和外衣）以及折射指数有关。

图 5-70 为光纤的衰减与波长的函数关系。由于瑞利（Rayleigh）散射，即 LWL 中材料的不均匀性所造成的散射，光纤衰减有个最低极限。这是物理上的，并不是技术上的极限。之后应该尽可能地选择较大的波长。当波长为 1.4μm 时有一个衰减峰值，它是由玻璃中不可避免的 OH 离子引起的，因此这个高度与生产工艺有关。当波长为 1.6μm 时衰减值又继续升高，因为此时引起了 SiO_2 分子中原子的振动。现代的光波导几乎能到达理论上衰减的最小值，即波长为 1.5μm 时的 0.2dB/km。作为比较：通常的窗户玻璃的衰减值是 100dB/km。

人们使用 3 个波长范围（光窗口）来实现光信号传输：0.85μm、1.30μm 和 1.55μm。每个范围都有不可想象的数 THz 的带宽！

通过一根光纤，理论上就能将人类所掌握的所有知识在 20s 内传输完毕。

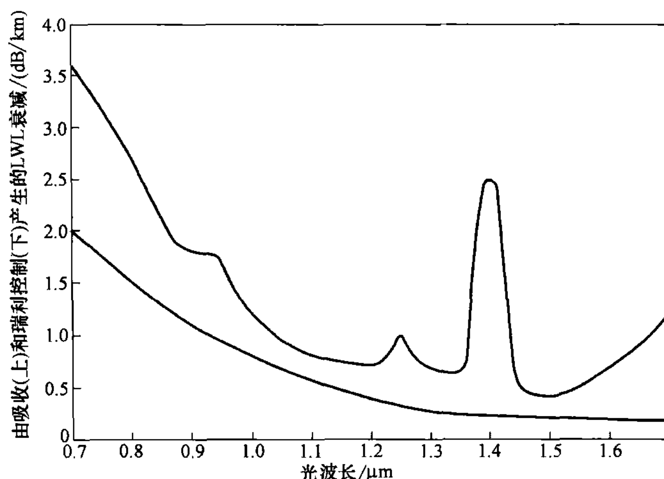


图 5-70 光波导的衰减特性曲线（解释在正文中）

0.85μm（接近红外线）范围的优点是电和光电元器件使用相同的材料，即砷化镓就能够实现。目前，考虑到较小的衰减，人们喜欢将窗口选择在较远的红外线范围，

此时 1.55μm 处的第三个窗口的缺点是转换器较贵。

对短距离应用（例如测量技术）而言不用考虑衰减，因此也可以使用可见光。有时在这些应用中也用较粗但比较便宜的人造材料 LWL 来代替。

除了衰减，还会遇到其他作用，即不同类型的离差（运行失真），其中有两个比较重要：

1）材料离差（也被称为色彩离差）：运行时间与波长有关。

2）模式离差：经过不同路径的辐射有不同的运行时间，例如图 5-69 中的辐射 A 和辐射 B。

运行时间的不同会导致码间干扰（ISI），LWL 越长，ISI 越大。如果不用特殊的方法来消除这些离差，离差（而不是衰减）则是限制转发器之间最大距离的标准。因此 LWL 的能力不是用 Bit/s 给出，而是用 Bit/s 乘以 km 来给出。例如，一个给定的 LWL 可以在 100km 的距离内传输 100MBit/s，或在 10km 的距离内传输 1GBit/s。

通过使用激光二极管（LD）来代替便宜的发光二极管（LED）就能够减小材料离差，因为前者能发出窄带光信号。

图 5-69 所示的光纤是所谓的阶梯剖面光纤，因为折射指数在从内核到外衣的过渡中在跳跃性地减小。如果折射指数的变化是连续的（坡度剖面光纤），或者是内核做得很薄（5μm），以至于只能存在一个模式（图 5-69 中的辐射 A）（单模光纤），那么模式离差就会减小。单模光纤也有一个阶梯剖面。坡度剖面光纤和阶梯剖面光纤比单模光纤要粗很多，它们的内核直径为 50μm，外衣直径为 125μm。

粗的阶梯剖面光纤只有说教的价值。单模光纤从距离上来说能提供最大的传输能力。但是它的孔径很小，即光线的入耦合并不容易。由于内核很细（5μm，与光波长 1.3μm 或 1.55μm 有相同的数量级），插接、接合等都比较难实现。尽管如此，在需要高速数据传输速率并且距离较远的地方都使用了单模光纤。在不久的未来人们只使用单模光纤，因为坡度光纤价格上的优势正在逐渐消失。

LWL 对湿度的入侵、机械上的压力（如牵拉、较大的弯曲、侧面的压力）等都是很敏感的。因此 LWL 的生产实际上就是包装技术！

不同的 LWL 段通过接合永远连接在一起，通过插接可暂时连接在一起。接合和插接使用相应的辅助工具来完成。比较困难的是光纤和转换器的连接。通常二极管的生产厂家通过在出厂时就将一小段 LWL 绑在半导体上的方法，来帮助完成这项任务。引入的衰减，在接合时要按照 0.1～0.5dB 来计算，在插接时按照 0.5～1dB 来计算。

在经常需要插接的插口上有时使用透镜将光集中并接在另一头。接触面的一个脏点只会使衰减升高，并不影响接触。

3. 电光转换器

作为发送器件目前只有两个组件可供选择：发光二极管（LED）和激光二极管（LD）。表 5-10 将这两个器件的特性列在了一起，+ 和 - 表示哪个胜出或转差。

在使用 LD 和 LED 时人们通过调节前向电流的变化来调节光的强度。此时这两个组件表现出不同的特性：

表 5-10 LED（发光二极管）和 LD（激光二极管）的对比

LED	标准，特性	LD
-	光功率	+
-	频谱宽度（材料散射性）	+
-	与光纤的耦合（较小的空间辐射瓣）	+
-	调制带宽	+
+	电路复杂性	-
+	生存时间	-
+	对研究的危险性（很强的不可见 IR 辐射！）	-
+	价格	-

1）LED：调制特性曲线是非线性的，这在模拟信号传输时会产生失真（失真系数）。如果信息信号的带宽不是很大，那么可以先实现一个 RFM、PFM 这种形式（见图 3-36）的预调制，将信息放在信号的时间轴上。

2）LD：特性曲线出现弯曲。超过这个膨胀才开始真正的激光运行（受激励的发射），此时特性曲线几乎是线性的。因此，甚至分布式有线电视网中使用的模拟 FDM 的传输都是可行的。

作为接收器件，同样也只有两种组件需要考虑，即光敏二极管（PIN-Diode）和雪崩光敏二极管（Avalanche Photo Diode，APD）。这两个二级管都工作在阻断的方向上，反向阻断电流随阻断区内穿过的光量成函数变化。因此在 p 半导体和 n 半导体之间的阻断区随着固有 i 层而变宽，这也就是 PIN-diode 这个名称的由来。在 APD 的情况下，除此之外还会出现大量的载荷子，这是由较高的阻断电压产生的。使用一个跨导放大器，可将与信号有关的阻断电流转换成电压。表 5-11 比较了这两个接收器件。

表 5-11 PIN 二极管和 APD 的对比

PIN	标准，特性	APD	PIN	标准，特性	APD
-	敏感性	+	+	固有噪声	-
-	带宽	+	+	电路复杂性	-

很明显，APD 将要被 PIN-Diode 所代替，因为后者的噪声较小，并且光功率并不是一个限制标准。系统的噪声和系统的线性由发送方来确定。

4. 光线路

根据表 5-10 和表 5-11 可以得出下面的结论：

1）LD 主要用在长距离和高速率传输中，最好与单模光纤结合起来使用。

2）LED 单模光纤的组合是没有意义的，因为此时很小的光功率只能被很差地耦合进去。

表 5-12 给出了一些有意义的发送器、光纤和接收器的组合。这里有意识地没有特别指明窄带、宽带，以及长和短等表达，因为技术上的极限值总是在不断变化的。对于远距离传输和有线电视（CATV）只使用单模光纤。部分 LAN 还使用坡度光纤。

表 5-12 光传输有意义的单元组合

	窄带信号	宽带信号
短距离	LED 分级索引光纤 PIN	LD 分级索引光纤 PIN/APD
长距离	LD 分级索引光纤 PIN/APD	LD 单模光纤 PIN/APD

光传输时的功率投入与微波链路的情况，即式（5-22）是相同的。只是用光纤的衰减来代替自由空间衰减，并且增加已经提到过的接合和插接衰减。当然，天线增益在光纤传输中是没有的，相反有一个进入 LWL 的光入耦合的有效系数。

5. 调制和复用

与光窗口的数 THz 的带宽相比，目前信息信号的带宽只是很可笑的。从 LWL 的角度来说光传输是显然的窄带传输，尽管人们一直在说宽带传输。关键是相对带宽，即信号的带宽与载频之比。在 200THz 的载频上，所有技术上正常的内容都是窄带的。

（到目前为止）光只能在它的强度上进行调制。人们称之为 IM，这同 AM 等具有相同的含义。这个调制可以是模拟的（连续的）或是数字的（离散的）。

在模拟调制时，必须考虑失真系数和温度漂移，这主要是针对发送二极管的。因此人们喜欢使用已经提到过的预调制（Subcarrier-Modulation）。

数字调制方法使用光脉冲和时分复用（TDM）。在发送二极管的输入端以及接收二极管的输出端是正常的、像在第 2 章中描述的那样的基带信号。这种强度调制是一种光的 ASK。LED 时人们令调制深度 m 为 1（On-Off-Keying，OOK）。相反对激光二极管来说 $m \approx 0.9$，这样激光二极管可以一直保持在激光模式（当前向电流较弱时，激光二极管就像一个 LED）。

误码率主要是由高斯噪声引起的。这与无线连接刚好相反，那里通常情况下突发干扰是主要敌人。如果需要信道编码，那么在 LWL 系统中选择卷积码比较方便。

目前的系统使用 10GBit/s 的速率，40GBit/s 的速率只在实验室中使用。此时只使用一个波长的波。当同时使用多个波长（“颜色”）的波时，光纤会被更好地利用，人们将它称为 WDM（Wavelength Division Multiplex）。这同 FDM 是一样的，但在光学上人们更喜欢说波长而不是频率。每个载波的距离根据单个信息信号的带宽以及 ITU-T 的标准来确定，例如可以实现 16×10GHz 或 32×5GHz 的系统。只有使用激光二极管时 WDM 才是可行的，因为只有它才能提供必要的窄频谱。每个载频都使用 TDM 信号来调制。也有与 3.4.6 节所说的 OFDM 方法类似的方法。

从单载频系统过渡到 WDM 系统，只需要更换其中的电子设备即可，光纤可以继续使用。

使用一个标准的单模光纤目前每个载频使用 10GBit/s 可以传输 100km 的距离。长距离可以使用转发器来实现，但它们还工作在电子领域，每 50～100km 就需要一个。对于特殊情况，例如跨越海洋，已经有了光放大器。

光放大器基于掺杂质的铒光纤放大器（Erbium-Doped Fibre Amplifier，EDFA），物理上能提供的带宽是 5THz，即只有光纤带宽的 10%。幸运的是 EDFA 工作在只有很小衰减的第三个光窗口上。为了弥补“较窄的”带宽，人们试图尽可能地减小 WDM 信号的每个载频之间的距离，即在 5THz 上提高载频的数量。如果载频的距离是 100GHz 或更小，人们称之为 DWDM（Dense WDM）。

目前在实验室中可以达到用一根光纤以 10TBit/s 的速率传输 100km 的距离。此时有 256 个 DWDM 信道，每个信道的速率是 40GBit/s。如果额外使用 7% 的开销给 FEC 信道编码，那么 10^{-4} 的误码率可以再提高一个数量级。

多数情况下人们使用两个相互分离的光纤来实现双向连接。因为包装是 LWL 最复杂的部分，所以通常许多根光纤共用一个外包装。多数情况下会直接提供光纤束来给人们使用。

只使用一根光纤也能实现双向连接。方向的分离使用光滤波器来实现，即一个方向使用 1.30μm，而另一方向使用例如 1.55μm 的波（FDD）。通过一个半透明的镜子可以设置一个颜色滤波器，这样也能应用乒乓系统。

6. 未来展望

由瑞利（Rayleigh）散射所决定的光纤衰减的理论极限几乎已经到达了，参见图 5-70。在实践中衰减已经达到了 0.22dB/km，因为可能的光纤最大长度是由离差决定的，所以人们正在这个方向上进行研究。例如人们在寻找特殊的脉冲形状，即所谓的孤立子，它对离差具有免疫力，因此转发器的距离可以提高到几千千米。

在实验室中人们继续研制光器件，如转发器、开关、混频器等，最终目标是实现全光网络，即连续一致的光传输，并结合光交换等。WDM 技术是这样的网络的一个前提条件，它还能在数据速率上带来巨大的突破。

5.6.4 本地环路的新概念

在几十年长的时间里人们对电信的理解几乎仅仅是通过电话网的话音传输。为此家里或办公室通过铜线被连接起来，使用复用和交换的远距离传输可以比较便宜地实现，如图 1-49 所示。

与此同时，用户的需求在不断增加，他们需要更多的业务。为此对传输网络的要求也发生了变化，参见表 5-13。这里还没有强调是一个移动用户还是一个静止用户要求这些业务的。

表 5-13 不同的电信业务

业务	应用实例	技术要求
话音	电话	低速率、双工、实时
无线广播	广播/电视	宽带、单向、实时
数据	E-mail	低速率，非实时
	文件传输	可变速率，非实时
多媒体	有音频和视频应用的互联网	一个方向具有较高速率，反方向速率可以较低，根据应用实时
双向多媒体	视频会议	两个方向都为高速率，实时

在网络领域内，对数据传输速率爆炸性的需求，利用光纤技术、卫星技术、复用技术以及交换技术，在可支付得起的费用范围内都能够得到满足。但在网络接入处，即本地环路，情况却是另外一种样子。在有线传输中，费用只有一小部分是给了媒体本身。大部分费用都用在线路的安装上，即挖掘、焊接、拼接及支持工作。因此电话网中 2/3 的费用都用在用户范围内的细小分枝上。

在目前没有管制的电信市场上，以前的垄断公司又要面对新的竞争：他们既不拥有本地环路，因为费用的关系他们也不能再安装第二套本地环路。因此他们很痛苦地与人协商共同使用本地环路的费用问题。

本地环路（last mile）的铜线是为最大传输距离为 8km 的基带话音传输而考虑的。如果不做其他事情，这些线路还不能提供具有一定数据传输速率要求的数字业务。下面这些概念会提供一些帮助：

1）调制解调器（Modems）：图 1-49 所示的接入交换机对信号进行数字化，这样能够使用 TDM 比较便宜地继续传输。基于这个考虑，使用反混淆滤波器把信号限制到标准的带宽为 3.1kHz 的话音信道上，如图 1-53 所示。使用调制解调器，人们可以利用这个 BP 信道进行数字传输，目前设备的速率是 56kBit/s。因为波特率是受限的，这些调制解调器必须使用像 QAM 那样的多进制调制。

2）ISDN（参见 6.4.2 节）：数字化的工作被放在了终端设备上。接入交换机的电子电路将由数字化的用户接入交换机来代替。因此本地环路的带宽不受反混淆滤波器的限制，而是受铜线电缆本身的频率的限制。ISDN 的传输速率是 144kBit/s。

3）HDSL/ADSL：这些缩写的含义是 High Speed Digital Subscribe Line（高速数字用户线）以及 Asynmetrical Digital Subscribe Line（非对称数字用户线）。它涉及的是很巧妙的调制方法，它能更好地利用旧的铜线电缆的信道能力。HDSL 在几千米长的距离上不使用转发器可以实现 2MBit/s 的双向传输，而 ADSL 在一个方向上可以实现 8MBit/s，在另一个方向上使用 128kBit/s。关于这点后面还有详述。

4）CATV：通过给每个房间布置同轴电缆，消除了每座房子的天线转换器。由于电缆的容量较大，能够给用户提供多达 60 套电视节目和无数的广播节目。现在人们代替电视节目而使用这条电缆来提供数据业务，反方向则通过正常的电话网或同样使用 CATV 网络。

5）LWL：一种很舒服但几乎支付不起费用的解决方案是，在每个房间和每间办公室都铺设光纤（FITL，Fibre In The Loop 或 FTTH，Fibre To The Home）。一种辅助方案是 FTTB（Fibre To The Building）和 FTTC（Fibre To The Curb）。

6）PLC（PowerLine Communication）：通过供电线来输出数据：原则上强电线构成了第二张分布很广的铜线网，虽然要放置许多转换器，但它可以提供几 MBit/s 的数据传输。

7）无线本地环路：用无线来连接用户（Radio In The Loop，RITL）比铺设新的电缆要便宜得多。但另一方面，现在的无线频谱资源已经很紧张了，并且还存在着电子烟尘问题。因此无线解决方案必须基于蜂窝无线技术（参见 6.5.2 节）并且工作在微蜂窝上，这样所用的发送功率会比较小。

对于 HDSL 来说，有两种方法可用，即使用 2B1Q 线路编码的基带传输和使用 CAP（无载波的 AM/PM，相当于 APSK 或 QAM）的数字调制。难点在于随着频率的升高而增大的线路衰减、串扰和回声。因此两种方法都需要均衡器和信道编码。

ADSL 同样有两种方法可用，即使用 CAP 和 DMT（Discrete Multi-Tone，这与 3.4.6 节所解释的 OFDM 方法没什么两样）。趋势是偏向于使用 DMT，频率范围是 1.1MHz，256 个载频，每个载频占用带宽 4.3kHz。225 个高频端的信道分给下行链路使用，使用 QAM 每个信道的速率为 35.5kBit/s。这有点类似于范围在 300 ~ 3.1kHz 的现代电话调制解调器。每个子载波 QAM 调制的进制数与衰减和反射相适配，是可调的。上行链路使用低频端的信道，这里最多还有 20kHz 的范围可用。使用无源滤波器能将这个低频范围分割开，可供正常的电话使用（Plain Old Telephone Service，POTS）。由于有这个

无源滤波器，当 ADSL 部分由于损坏而瘫痪时，POTS 部分还能正常工作。对于信道编码来说，通常使用 Reed-Solomon-Code，也可以选用交织作为补充。

因为通信技术还处于快速的发展过程中，不断会有新产品出现，人们也还需要新的名字。总的来说人们将这种能够对本地环路充分利用的、有效的数字方法称为 xDSL。这里 x 可以用 H（high speed）、A（asymmetrical）、S（summetrical）、V（very high speed）或 RA（rate adaptive）来代替。使用 VHDSL 可以在速率为 34MBit/s 的情况下传输超过约 2km 的距离。这个速率相当于 ITU-T 标准的 E3 复用层次，而 HDSL 可以传输 E1 的复用，参见 6.3.2 节。

xDSL 和 PLC 方法并不是没有争议的。因为传输媒体是无屏蔽的线路，加载了在 MBit/s 这个范围的数字信号，在整个的长波、中波和短波范围内会导致有损害的干扰雾。这会严重损坏在这个范围内的无线接收。

5.6.5 小结

图 5-71 所示为总结在一起的有线传输媒介。这里所给出的数据传输速率和距离只是考虑值。在实验实里这些值早就被超过了。

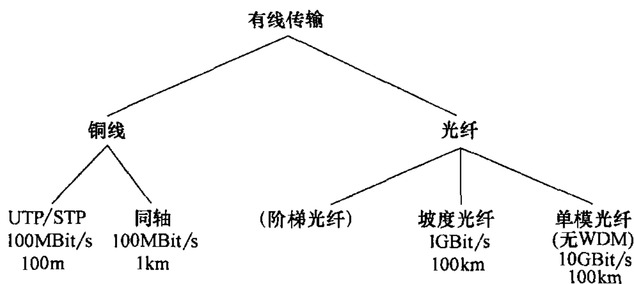


图 5-71 有线传输总结

最后还有个思考游戏，将上面有印象的数据以另一种方式关联起来。这个想法来自参考文献【Tan97】。

当信息以磁的方式存储，并要传到另一个存储器中时，数据传输的能力有多强？

我们观察一下长度为 240min 的录像带。根据表 1-4，数据传输速率的值为 216MBit/s。为了安全起见，我们用来计算的带宽为 5MHz，信噪比只有 40dB，根据式 (1-18)，数据传输速率为 66.7MBit/s。这样，一个录像带的存储约为 1TBit。

录像带的尺寸约为 19cm × 10.5cm × 3cm。大小为 40cm × 22cm × 30cm 的盒子可以装下 40 个录像带，5 个这样大小的盒子可以轻松地装在一辆小汽车上。这 200 个录像带可以存储 200TBit 的数据。

现在将这些盒子用 2h 的时间运输 100km 的距离。这样算下来数据传输速率约为 28GBit/s。使用单模光纤只要将电路一合就能完成了！

> 人们对电信技术所做的所有努力还只是满足次要目标，即提高数据传输速率。但最好的应该是减小时延！

第6章 通 信 网

本章的引言其实已在1.1.10节中开始了。这里还需要说明的是，通信网的发展本身就是一个独立的学科。本书80%的内容是关于传输技术的，20%是关于网络的。其实相反的比重也是可以的——本章只是通信网的简单概述，更多的内容请参见参考文献【Pet00】、【Sta97】及【Tan97】等。

6.1 OSI模型

通信网有许多完全不同的任务需要完成，这可能会导致3种后果：

1）只见树木，不见森林。

2）人们互相通话，但对通话方的内容却不了解。

3）不同设备制造商的设备不能互通。

鉴于存在这些问题将网络的复杂性进行分解，说一种大家都能理解的共同的语言是很有必要的。OSI模型的目的正是这个。这正如1.1.5节所介绍的：网络系统被划分为子系统。这个子系统用OSI模型的语言来说叫做层（Layer)。人们也管它叫ISO OSI模型，ISO模型或其他可想象的组合。参考模型也是常用的描述之一。

OSI模型的思想可以用参考文献【Tan97】中一个例子来描述。非洲和印度的两个哲学家想在家里就生活的意义进行辩论。他们找了两个翻译，每人都掌握本国的语言及英语，及两个话机。图6-1描述了这个连接的框图。

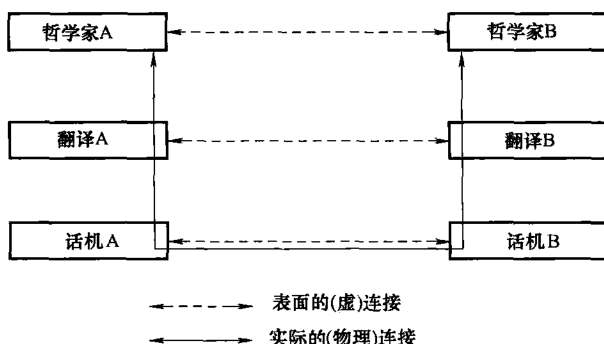


图6-1 OSI模型的简化示意图

两个哲学家相互进行内容上的通信（虚连接），物理上他们实际与他们的翻译在说话。翻译实际上也是这样，他们实际上在与话机进行交谈。物理上的连接是垂直的，除了最底层之外，它才真正连接到了远端。逻辑的连接一直是水平的，每人都在与自己的对方通话。打电话时人们虽然在对着传声器说，但却不是与传声器交谈。

分层模型最重要的是，每一层只需要理解它的对等层，并且与它邻接的上、下层有清晰的定义并遵守这个定义即可。

翻译必须说一种共同的语言，但是他们允许与自己方哲学家或是话机不用英语而使用法语。同样发话器也允许根据一定的规则从短波连接向卫星链路转换。

因此，人们通常将某一层从它相邻的下层得到的或是向它的上层提供的服务叫做业务。例如图 6-1 中最底层提供远距离传输的业务，中间层利用这种业务传输英语，而最高层基于此可以提供传输本国语言的业务。

OSI 模型包含 7 层，如图 6-2 所示。每一层的实体与通信对端位于同一层的实体使用协议进行通信。一个连接所使用的所有协议称为协议栈。除最底层之外，其他所有层均能用软件来实现，协议栈无非也是软件的集合。在一个具体的网络中，如果某层的功能不是必需的，这层的功能也可以省略掉。

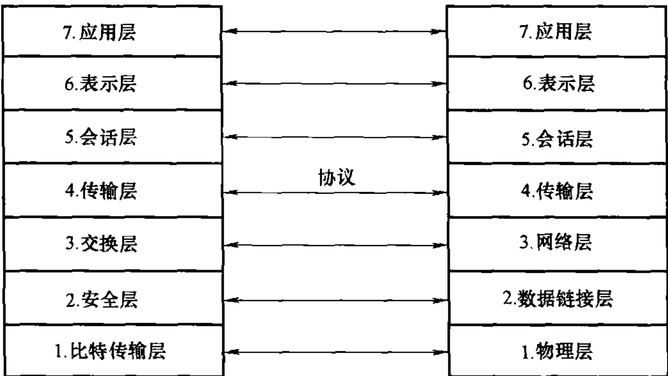


图 6-2 OSI 模型的 7 层

物理层属于传统的电子工程领域，电子工程领域的研究工作从底层向上展开。而应用层属于计算机领域的研究范畴，它自顶层向下展开研究。两者在中间层产生交叉。

物理层负责真正的传输，它负责建立从 A 到 B 的不安全的连接。在这层上必须掌握调制、载波、传输率、符号率等，具体的传输媒介（同轴电缆、光纤等）并不属于 7 层的范畴。本书的第 2、3 章及第 5 章已经详细介绍了物理层的知识，更详细的介绍是没有必要的。

数据链路层完成下面的多个任务：

1）将物理层不安全的传输转换成一个安全的传输。这依靠信道编码来实现，本书的 4.3 节已经作了介绍。经常使用的是 CRC 编码与 ARQ（Automatic Repeat Request，自动重复请求）协议相结合的方式。信道编码采用点到点之间的逐段方式进行，这种安全机制属于第 2 层。除信道编码之外还有端到端的安全机制，这属于传输层的范围。

2）流量控制：必须避免出现这样的情况，发送端发包很快，但接收端很慢，这样就会造成数据包的丢失。这通常与 ARQ 协议相结合来实现。在数据网中数据通常都是以分组的方式发送的，并且每个分组被接收后都要有收据。在一个拥有许多节点的网络中会导致这样的现象发生：因为发送端需要等前面分组被接收到的收据，两个分组

之间的时间间隔变得很大。好一点的情况是使用所谓的“管道”，在这种情况下发送端一次可以发送例如10个未被证实的分组。根据协议的配置，发送端在检测到分组丢失后可以从发生丢失的地方重传（go back n-ARQ，即ARQ中退回到n）或仅仅重传丢失的分组（选择性重传，即selective repeat）。在第二种情况下，接收端需要缓存，以便以正确的顺序将这些分组提交给上一层。

3）多媒体接入控制（media access，MAC子层）：在广播网络中有一个信道对所有的用户来说都是可以接入的，这会产生冲突。MAC子层负责处理这种情况（参见6.2.1节）。

网络层负责通过建立多个点到点连接从而形成整个网络中的连接，具体的内容包括交换（参见6.3.1节）、复用（参见6.3.2节）及寻路（routing，即路由）（参见6.3.3节）。

传输层是第一个涉及到端到端连接的实体。它包括连接的建立、数据的分段、端到端的差错控制、端到端的流量控制等。在从网络层到传输层的过渡过程中，状态由网络承载向网络可用转变。传输层的任务是向应用层提供易于使用的、有效的、可靠的、费用低廉的业务，同时使应用层与低层的实现细节和不足的地方分离开。如果存在多个备选网络的话，由传输层负责选择用于传输的网络。如果数据包在传输的过程中顺序发生改变的话，由传输层负责将数据包重新排序（采用数据报方式的分组交换时这种现象很容易发生，参见6.3.1节）。

传输层可以将一个网络连接（逻辑连接）分成多个物理连接来实现，这样可以提高整个连接的吞吐量。这属于图1-38中所描述的向下复用（Down Multiplexen）的一种情况。相反，传输层也可以使多个并行的逻辑连接复用成一个物理连接，以便节省费用。通常使用一个虚电路（参见6.3.1节），即不仅按传输的字节，也按使用虚电路的个数来计算费用，因为虚电路的个数将存储空间资源与路由资源联系在一起。

会话层打开、管理并结束一个会话。

表示层负责语法和格式方面的适配，这也涉及到源编码（参见4.1节）、端到端的加密（参见4.2节）等内容。

应用层负责识别用户、选择通信参数（服务质量）以及提供一些最基本的业务，如文件传输（File Transfer）、电子邮件（E-mail）等。

下面这个关于电话连接的例子能清晰地看出7层模型中不同层之间通信的过程。括号中给出的是上面介绍的各层。从一个可用的话机（物理层）上拿起受话器，就可以听到拨号音（数据链路层）。然后开始拨号（网络层），某个公司的电话运营商负责接收号码（传输层）。XY是要找的某人，并且以一种语言回答（“Do you speak English”，你讲英语吗？）（表示层）。这以后才能真正地说其他事情（应用层）。

图6-3描述的是从A到B通过图2-18介绍的转发器建立连接时OSI模型的使用情况，这种转发器放在图1-49描述的节点之间。

转发器还将会影响信道编码，也就是说使用点到点的差错安全机制代替端到端的

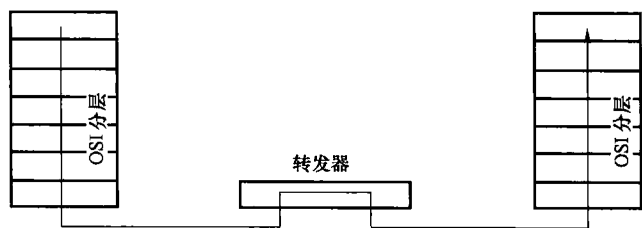


图 6-3 OSI 模型中的通过转发器的连接

安全机制，所以说中间的站点也涉及到第 2 层的内容。当不同的 LAN 需要连接时也涉及到这一层，这样的中间节点被称为网桥，如图 6-4 所示。

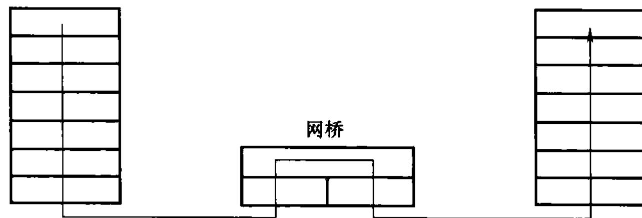


图 6-4 OSI 模型中的通过网桥的连接

如果中间站点还需要有路由功能的话，它就还需要第 3 层的功能，这时的中间站点被称为路由器，如图 6-5 所示。图 1-49 所描述的节点就是这样的路由器。

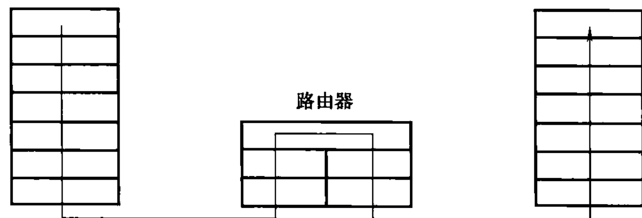


图 6-5 OSI 模型中的通过路由器的连接

网关是这样的一个中间站点，它除了路由功能之外还有协议转换的功能。不同的网络之间通过网关来连接，因此它还涉及到比图 6-5 所描述的更多的层次。

前面已经介绍过了，每一层都使用它下一层提供的业务，并向它的上一层提供更好的业务。这个提供点（接口）被称为 SAP（Service Access Point，业务接入点）。如果某一层还需对它的业务进行完善，还需要资源，在这种情况下就是带宽。带宽还被用在根据不同的层次所带来的不同开销，即一些多余的比特。这些开销对于不同层次上的协议产生作用是必需的。图 6-6 描述的是从第 4 层开始向下分组的构建过程，向上分组的构建过程原理相同。

第 3 层的开销包括地址信息和帧复用信息，第 2 层的开销包括信道编码需要的冗余，第 1 层的开销包括接收端产生时钟信号需要的同步信息和接收端重新产生信号需要的训练序列。在实际中这些开销可以被分成两段（Header and Trailer，头），这仅仅

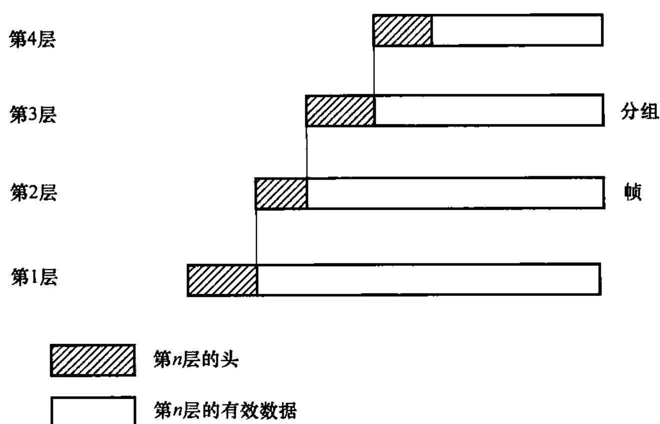


图 6-6 每一层为了完成任务需要一定的开销

是一个规范的问题。除此之外，每一层还能根据自己的需要产生自己层的分组包，如第 3 层可以为了寻找最优路由信息而发送路由发现的分组包。各层的总开销可能会达到第 1 层传输的数据总量的 50%。

网络的发展与没有控制的各种标准的增长分不开，因为各大公司都试图将他们自己的开发成果变成标准。很明显，为了实现不同的应用目的需要使用不同的网络技术。其实很多子问题都可以采用统一的方法解决，但遗憾的是很多机会都错过了。OSI 模型是个尝试，至少能发现一个共同的语言。

OSI 模型虽然曾遭到很大的批评，但却广泛地应用起来。这个模型的最大问题是它被通信网络分段。正如在局域网（LAN）中的使用一样，广播网络的部分任务不太容易构建在 OSI 模型中的各层上。这导致了在 LAN 中二层负担过重，而三层却被动地无所事事。关于第 5 层人们还不知道如何开始使用。虽然有些批评是合理的，但 OSI 模型还是很有价值的，至少它能让各种网元使用共同的语言。TCP/IP 参考模型（Transmission Control Protocol/Internet Protocol，传输控制协议/互联网协议）与 OSI 模型存在竞争关系，它只涉及 4 个层次，如图 6-7 所示。这里所说的互联网并不是指通常意义

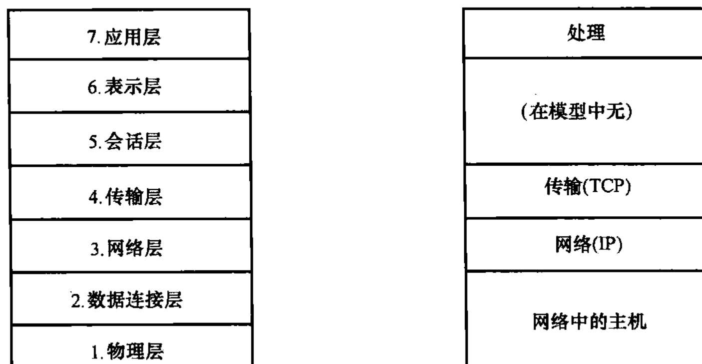


图 6-7 OSI 模型（左）与 TCP/IP 模型（右）的比较

上的网络，它仅仅是模型中的一个层次。

传输层比较有名的协议是 TCP（Transmission Control Protocol，传输控制协议），通常与 IP（互联网协议）结合起来使用而被记为 TCP/IP，它控制了整个 UNIX 世界。TCP/IP 最初是美国国防部需求的一种方法。基于此（不仅仅是因为它无可竞争的质量保证）该协议栈得到了广泛的推广。

6.2 LAN 及 MAN 技术

在 LAN 和 MAN 中，第 2 层的负担很重。因此人们将它分成了两个子层，即负责安全和流量控制的 LLC（Logical Link Control，逻辑链路控制）子层和负责多接入控制的 MAC（Media Access Control，媒体接入控制）子层。

6.2.1 多接入控制（MAC 子层）

多接入控制仅对广播网络很有意义，这种网络仅有唯一的一个连接所有用户的物理信道。局域网，如以太网或是令牌环网，都按照这种原理工作；无线网络，特别是卫星网络，也都属于广播网络。人们围绕在一起讨论也可以看成是一个广播网络，因为所有人都使用一个通道讲话，但却不允许同时使用。为了避免发生冲突，需要有一种规定。例如一种常用的规定是采用协调人的策略，有时也可以强制地根据音量来决定谁来使用信道。

1. FDMA：在频域内分离

FMDA 的含义是 Frequency Division Multiple Access（频分多址），它意味着一个信道在频域内被多个用户所瓜分。因此，FDMA 与 FDM（Frequency Division Multiplex，频分复用）类似。

2. TDMA：在时域内分离

TDMA（Time Division Multiple Access，时分多址）指相互竞争的信道使用者按照时隙被分开来。这需要在两个站点之间进行时钟同步，可是随着网络范围的扩大，如卫星网络，这一点不太容易实现。但另一方面，时隙的切换比频域的滤波容易，因此它的效率比 FDMA 要高。

FDMA 和 TDMA 的缺点是没被使用的频率或时隙不能被其他用户所使用，也就是说信道没有被很好地利用。FDM 和 TDM 这两种复用方法却可以通过交换将这个缺点避免。MAC 所研究的问题也能解决这个问题，但它却需要另外一个信道。对人们围在一起讨论这种情况来说，一个想要发言的人可以通过举手引起其他人的注意。企业无线网络（参见 6.5.3 节）使用 TDMA 或 FDMA，及一个特殊的工作信道。

3. CDMA：在码域内分离

CDMA（Code Division Multiple Access）同时给所有的用户所有的带宽，它在所有 3 个轴上对香农立方体进行划分，如图 1-36 所示。在具体实现上，它使用扩频调制（参见 5.3 节），每个用户都有一个扩频序列，这个扩频序列与网络中使用的其他扩频序列是正交的。用户的分离就在于分配序列，这是在运营之前的纯组织工作。运营时，每

个用户都可以随时开始使用网络，只要网络的能力还没被耗尽。正是这种在运营时不同的连接之间不需要同步的特性，使 CDMA 对卫星系统有很大的吸引力。

作为一个比喻，大家可以想象一下有个鸡尾酒会，此时成对的人开始讨论[Tan97]。在 TDMA 的情况下，大家都站在房间的中央，一个接一个地开始讲话。在 FDMA 的情况下，每对参加讨论的人被分在不同的房间，大家可以同时讲话。在 CDMA 的情况下，大家都站在房间的中央并同时讲话，只是一对说德语，一对说英语等。此时外语全当是背景噪声。

4. CSMA/CD：通过竞争来分离

一种经常被使用的方法是 CSMA/CD（Carrier Sense Multiple Access with Collision Detection)。此时每个想要发送数据的站点不断地监听信道，只要信道空闲就占用它(carrier sense)。因为在信道中传播有一定的时延，因此会出现多个站点同时开始发送的现象，此时数据总线上的数据出现叠加因而被损坏。这些冲突能够被站点发现（collision detection)，因此停止发送。现在所有的站点都开始等待信道再次变成空闲，并开始下一轮的碰撞。为了避免这种现象发生，每个站点都有一个随机计时器，当信道变成空闲时启动它。如果当计时器超时时信道仍为空闲，那么站点就可以开始发送。

在无线网络中，冲突检测是不可能的，因为自己的发送信号淹没了所有其他远处的信号，因此那里只能应用 CSMA。冲突只能根据所发出的数据包没有被证实来确定，之后这个站点会重新试着发送。另外一个计数器（retry counter）监视未成功的连接，当重试的次数过多时，将站点从网络中断下来。这样可以避免在负载很大的网络中只有冲突出现，而没有成功的数据包传输。

5. 轮询/DAMA：通过协调来分离

只要没有过多的站点来争抢信道，即发生冲突的包所占的比例较低，CSMA 会工作得很好。当信道的负载较大时，它的有效性就急剧下降。

在这种情况下，轮询是一种比较好的方法。此时会规定一个协调人，他轮流询问每个站点，是否有数据要发。这种方法的优点是能够确保在一定的时间间隔内询问到每个站点，而 CSMA 却可能在最终能够发送数据之前，持续等任意长的时间。这个缺点能够用 DAMA（Demand Assignment Multiple Access）来减轻。协调人还是询问每个站点是否有数据要发，但却留意是否有站点持续没有需求。如果有的话，后面对它的询问次数就会减少。在有中继的无线网络，协调人的任务当然由中继站点来承担。在网络中增加一个新的站点需要一个特殊的过程，但站点的离开能够很容易被协调人发现。

6. 令牌传递：通过预留来分离

令牌是能够发送数据包的权利。这些令牌是特殊的位图，站点能够自己分发。为此所有的站点必须组成一个逻辑上的环，即每个站点都要知道它后面是谁。物理上这些站点不一定组成环形结构。令牌会被不断地传递下去，想要发送数据的站点拿到它后，在上面做个占用标记，并将它要发送的数据附在上面，然后将它继续传递。后面的站点检查一下地址，如果发现数据包不是给它的，就将它继续传递。而接收者将这个数据包复制下来，然后在包的头部写上接收证实，同样将这个包继续传递。最终这个包返回到发送者，发送者识别出接收证实后，可以继续控制这些数据，并将令牌

释放。

当然会有特殊的方法将一个站点新加入到一个环中或将一个站点从环中去除。此外，令牌也可能被丢失，例如，当一个站点刚刚拥有令牌就出现问题而瘫痪了，这样在剩下的网络中必须能够重新生成一个新的令牌。

与轮询相同，令牌传递在信道负载较小时等待时间可能会毫无必要的很长，为此，可以给每个用户一个最小数据传输速率和最大等待时间保证。

6.2.2 逻辑链路控制

LLC 子层在 MAC 子层之上，两者一起构成了 OSI 模型的第 2 层（数据链路层）。LLC 子层的功能是差错控制和流量控制。

一种远距离传输经常使用的 LLC 协议是 HDLC（High-level Data Link Control）协议。所有的 X.25 网络都使用这个协议，在 X.25 中它被称为 LAPB（Link Access Procedure, Balanced）。同样这个协议也被用在 ISDN 的控制信道中，在那里它的名字为 LAPD（Link Access Procedure D-channel）。图 6-8 给出了 HDLC 的帧结构。

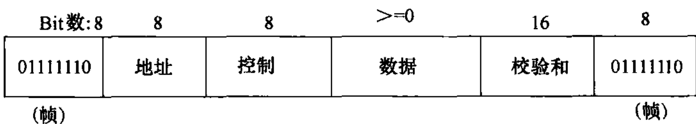


图 6-8 HDLC 帧

帧被分成多个域，它们的含义如下：

1）帧识别：比特序列清晰地标识帧的开始。之后数据的传输是透明的，即为了能够传输任意的比特组合，包括序列 01111110 在内，需要使用填充比特：从第 3 层发来的数据，每 5 个 1 之后，就要被填充一个 0。而接收器在每 5 个 1 之后就要将后面的 0 去掉。如果连续收到了 6 个 1，那么就可以放心地认为它是帧识别。

2）地址：表明这个帧属于一个特定的连接。

3）控制：这个域被用作帧编号、证实等，图 6-9 给出了它的具体结构。

4）数据：这里包括从第 3 层上发送过来的数据，必要时要进行分段（即被分成多个包）和填充。内容是任意的，长度也是自由的，不受 8Bit 结构的限制。

5）校验和：CRC 信道编码的校验和。

6）帧识别：由于有了填充，该域能唯一描述帧的结束。如果后面还有多个帧在传输，有一个 01111110 这样的字用来区分就够了。在第 1 层上可以实现线路编码。它将 01111110 这样的帧序列转换成适合接收器时钟再生的序列。但在这里第 2 层所涉及的只是逻辑状态而不是物理状态。

HDLC 协议有 3 种类型的帧，它们根据“控制”域（8Bit）来区分，如图 6-9 所示。

图 6-9 的含义为：

1）Seq：帧的编号。使用它可以有 7 个未经证实的帧，而且允许选择性重发。



图6-9 图6-8中控制域的构成

2）Next：对方站点对所收到的帧的证实域。这是一种“携带式（piggyback）”方法，即证实与数据在同一帧中传输。“Next”域包含3Bit的下一个期望的帧的号码。如果编号为010的帧被收到了，那么该域就填011。

3）P/F（poll/final）：使用该比特计算机可以对终端进行询问。如果计算机给终端发了一个P，那么终端就可以发送数据。如果有数据要发，仍将P比特置位，除非是最后一帧，它用F来标识。

4）Typ：帧的类型：

0 接收准备好：证实帧，如果因数据丢失，反方向上的证实使用“携带式”方法没有被收到，那么就要使用该帧。

1 拒绝：否定证实（使用“携带式”方法是不可能的）。

2 接收未准备好：肯定证实，但不希望继续收到帧，例如因为接收端没有足够的缓存（流量控制）。

3 选择性拒绝：要求重发某一指定的帧（selective repeat）。并不是所有的HDLC相关的协议都支持这种选择。

对于未编号帧，不同的HDLC版本会有较大的区别。使用“Typ”和“Modifier”两个域总共可以有32种可能性，但实际中并没有用到所有的情况。该域是作控制用的，例如：

1）SABM：Set Asynchronous Balanced Mode（连接建立）。

2）DISC：disconnect（从网络上终止连接）。

3）FRMR：frame reject（包结构出现验证错误）。

4）UA：Unnumbered Acknowledgement（对未编号帧的肯定证实，它们每个只能有一个可接受状态）。

6.2.3 LAN的分段

LAN是一种很典型的广播网络，因此当网上连接的站点数量很大时，网络效率不高。因此人们将计算机网络分成多个相互无关的LAN，然后用网桥将它们连在一起，如图6-10所示。通常给一个组织团体分配一个LAN，这样许多流量就会被限制在一定

的范围内，只有少数的流量才需要经过网桥到另一个 LAN 中。

网桥尽可能用硬件实现较高的吞吐量。在这种情况下人们称之为交换机，精确地说是二层交换机。

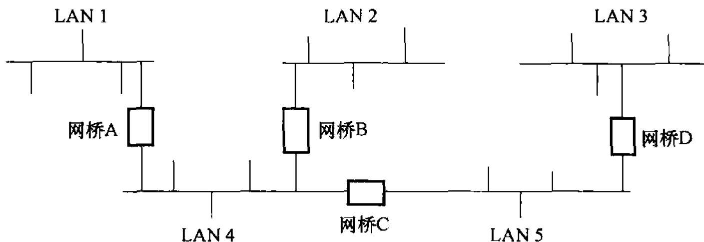


图 6-10 通过网桥来连接不同的 LAN

每个站点都有一个明确的 MAC 地址。例如 IEEE802.5 标准引入的地址长度是 16Bit。其中的 7Bit 用于识别 LAN，8Bit 用于识别站点，另 1Bit 用于区别组地址（组播）和一般的个体地址。目前实际中使用的 MAC 地址长度为 48Bit，其中的 14Bit 用于标识 LAN。

网桥使用 LAN 地址继续转发数据包，这是依靠转发表来完成的。转发表是自动生产的，在其他的网络配置时被更新。为了避免出现环路，人们构造了所谓的生成树，这是一颗树，能够明确地到达每一个网桥。如果在路径选择时有多种可能性，例如如果图 6-10 中在 LAN2 和 LAN3 之间还有另外一种耦合关系，那么根据开销函数可以在生成树中选择一条最优路径。

6.2.4 系统范例

在本节中我们讨论几种 LAN 的具体实现方式，这些网络从不同的公司发展而来，之后被 IEEE 标准化。图 6-11 给出了这些标准的概况。

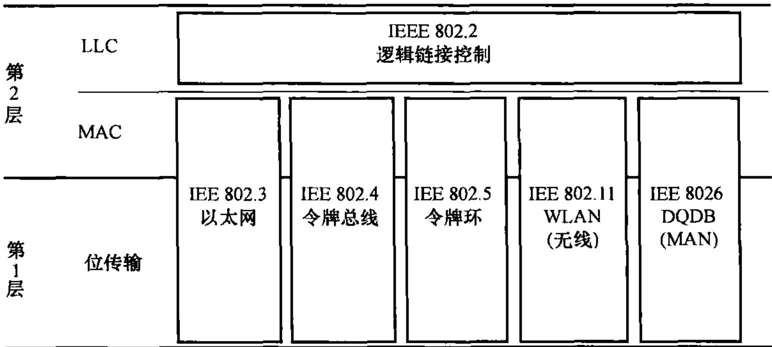


图 6-11 IEEE 针对 LAN 和 WAN 的标准工作

1. 以太网（Ethernet）、**快速以太网**（Fast Ethernet）、**吉比特以太网**（Giga-Ethernet）

以太网是最古老的但被最广泛使用的 LAN 技术，它约占市场的 60%。它有一个逻

辑总线，接入方法使用的是 CSMA/CD。Ether 称为以太，它的名字来源于与无线网络类似的工作方式。以太网中总线上的数据传输速率为 10MBit/s，快速以太网为 100MBit/s，而吉比特以太网的速率为 1GBit/s。

我们先来看一下最原始的、在 IEEE802.3 中标准化的以太网。传输媒体以前是较粗的同轴电缆，因为它的颜色也被简称为“黄电缆”。后来出现了较细的同轴电缆（细线）、UTP 电缆及光纤。表 6-1 将不同的媒体列在了一起。它们的名字是很容易辨认的：“10” 意味着 “10MBit/s”，“Base” 意思是它涉及的是基带传输，“5” 和 “2” 分别表示以 100m 为单位的段长，“T” 指双绞线，“F” 指光纤（fibre）。

表 6-1 以太网使用的电缆

名 称	类 型	段 长/m	说 明
10Base5	粗同轴电缆	500	过时的
10Base2	细同轴电缆	200	过时的
10Base-T	双绞线	100	
10Base-F	LWL	2000	

被传输的数据被分成帧（分组），图 6-12 给出了以太网的帧结构。物理传输是在基带上进行的。为此需要一个线路编码，这里选用了如图 2-4 所示的差分曼彻斯特码。

字节数: 7	1	2或6	2或6	2	0～1500	0～46	4
前导	帧	目标地址	源地址	长度	数据	填充	校验和

图 6-12 以太网帧的结构

以太网帧的各个域的含义和功能如下：

1）前导：7B 长，传输位图 10101010，以便于接收器的时钟再生器进行同步。

2）帧识别：用序列 10101011 来标识帧的开始，到它的最后一个 Bit 为止都属于前导。这是因为当没有数据流时，总线上会保持静止。接收器的时钟再生器必须首先进行同步，因此前导的开始部分会有很多错误。因此必须使用帧识别字作为共同的时间戳。

3）地址：通常情况下仅使用 6B 的地址。它也可以被设置成组地址，这样多个接收者可以有效地接收同一个包（组播）。全 1 的目标地址是指网络中的所有站点（广播）。

4）长度：用字节表示的数据域的长度。

5）数据：实际的有用数据。

6）填充：帧的总长度最小必须是 64B，这样发生冲突的时间才足够长，所有的站点才能够识别出冲突。如果帧太短的话，在数据到达网络中最远一个站点之前，发送者可能就停止发送数据。这样在那里可能会出现冲突，但却不能确定发送站点。因此如果数据域太短的话，就用填充域来相应地延长帧。

7）校验和：CRC 误差检测的控制字。

CSMA/CD 方法的缺点是当网络的负载很大时会出现网络崩溃的趋势。因为此时几乎只剩下冲突，而没有有用数据的传输。因此实际中会根据流量限制每个网段中站点的数量，这个数字远低于标准规定的 1024 个。为了提高以太网的效率，还引入了不同的策略：

1）空间上较大的地方在物理上用转发器来隔离开，但逻辑上和功能上都属于一个子段。

2）当连接的站点增多时，用交换机将 LAN 分成不同的接入范围（collision domains）。

3）用网桥将多个 LAN 耦合成骨干网，这比图 6-10 所示的无结构的耦合要好图中 LAN4 要负责 LAN1 中所有的需要跨不同边界的流量。

如果 LAN 相距较近，那么可使用路由器来代替骨干网，所有 LAN 都可连到路由器上（collapsed backbone）。

4）骨干网可以是另一个以太网（当然优选快速以太网或吉比特以太网），也可以是使用其他快速技术的网络（如 FDDI 或 ATM）。

5）逻辑上，每个以太网都能用总线来实现。物理上人们喜欢采用星形布线，并在星形交叉点放一个集线器（Hub），甚至是一个交换机。这就相当于图 5-67 所示的布线方式。

集线器由转发器构成，每个输入信号都被转发到所有的输出线上，实际上是模拟一个广播方式。相反，交换机可以避免冲突，它将广播网络分离成点到点的传输。

6）交换机能够识别出帧的源地址，并因此获悉哪个站点连的是哪个端口（Switch-Port），所以交换机具有将到来的帧只转发到希望的出口的能力。如果在交换机上再安装一个存储器，那么它可以同时转发与最开始的数据交换没有关系的站点之间的帧，人们将它称之为以太网交换。目前的趋势是实现全双向连接的交换。

7）如果中央服务器连在交换机上，那么这条线路上的负载就会很大，因此这条连接的能力要用比其他连接的能力都大很多。

快速以太网从技术上来说（布线、MAC 和帧格式）同以太网是相同的，只是它的速率可以提高到 100MBit/s。表 6-2 描述了所用的电缆。

表 6-2 快速以太网使用的电缆

名称	类型	段长/m	说明
10Base-T4	双绞线，UTP-3，4 对	100	很少使用
10Base-TX	双绞线，UTP-5，2 对	100	
10Base-FX	LWL(2 坡度光纤)	136	

以太网帧的最小长度为 64B = 512Bit。当速率为 10MBit/s 时，传输这么长的帧所用的时间为 50μs。一帧在这么短的时间内能以光速的 2/3 的传播速度（对所有的 PE 隔离的电缆来说这是一个很好的经验值）走 10km 的距离。

冲突能够被检测到的最晚时间点是，当站点 A 发出的帧，在到达站点 B 前的短短一

刹那，B开始发送。只要这一帧还没完全发完，A就能意识到冲突，因此帧的发送时间必须是从A到B的传播时间的两倍。在一个以太网中，使用Hub时离Hub的最远距离只能是上面所计算的10km的1/4。考虑到本身的传输时延，因此在10MBit/s的以太网中与Hub的距离被限制在2km之内。电缆的衰减会使这个值继续减小，参见表6-1。如果考虑到兼容的原因，在快速以太网中保留相同的帧结构，那么每段的长度会减小到200m。

另外一个问题是带宽。在100MBit/s的速率下使用UTP电缆，很占用带宽的曼彻斯特码是不合适的。使用100Base-T4，可以将它向下复用到3对线上，这样速率会减小到原来的1/3，第4对线用于报告冲突。除此之外还使用了8B6T线路编码（8个二进制Bit被变成6个三进制符号）。每对线上的符号率为 $(1/3)\times(6/8)\times(100)\times10^6=$ 25MBaud。根据表5-7，UTP-3电缆的带宽为16MHz，根据式（2-5）和式(2-3)，上面所计算的波特率是可以传输的。因为8B6T码与曼彻斯特码有明显的区别，要使用一个MII（Media Independent Interface）来进行适配。

最后我们还得提一下吉比特以太网，它的速率是1GBit/s，主要用在骨干网和服务器的连接上。根据上面的计算，最大段长必须减小到20m，这样帧结构才能保持不变。但这个距离太短了，因此人们不使用UTP电缆而是用LWL来代替。除此之外还使用双向连接，并增加了交换机。

2. 令牌总线、令牌环和FDDI

以太网的缺点是终端只能以统计的概率接入网络，以及最开始的标准（目前的长度是超出1200页）没有帧的优先级。这不能满足具有实时要求的应用，如生产线。对负载很大、要求相互间进行合作的CSMA网络，吞吐量下降得很快。这种现象在使用令牌的接入方法的情况下不会出现。

对于使用令牌总线的网络，各站点物理上以总线或树形结构连接，但逻辑上却组成一个环。这种逻辑结构的实现方法是每个站点都保存有它的前一个站点及后一个站点的地址，这样才能满足令牌传递的需求。这种情况下电缆本身是具有广播特性的介质。通常使用75Ω电阻的同轴电缆，信号采用FSK或PSK调制。数据传输速率可为1MBit/s，5MBit/s或10MBit/s。

令牌环网络中，各个站点物理上与逻辑上都是一个环。令牌环的优点是：

1）环不是广播，而是点到点连接的首尾相连。

2）所有站点具有相同的权重，其中的一个站点可以作为仲裁人。

3）以太网使用的模拟组件不用进行冲突检测。

令牌环使用差分曼彻斯特编码，使用双绞铜线时速率可为4MBit/s或16MBit/s。通常最多250个站点通过双线连接在一个线中心，这样可以桥接旁路中继损坏的环形扇面。图6-13所示为令牌环帧。

与6-12相比可以看出，令牌环帧与以太网帧有很大的相似性。在OSI第2层可以有多种变化，但它们在LLC子层已经进行统一，如图6-11所示。正是由于这个原因，令牌环网与以太网可以通过所谓的转换桥（Translation Bridge）进行连接，如图6-4所示。

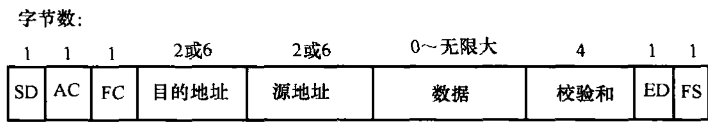


图 6-13 令牌环帧的组成

图 6-13 中各域的含义如下：

1）SD（Start Delimiter，始帧界）和 ED（End Delimiter，尾帧界）：帧的识别。

2）AC（Access Control，接入控制）：接入控制，包含令牌比特（Token Bit）、优先级、保留比特及监视比特。

3）FC（Frame Control，帧控制）：帧控制，用来区别数据帧和其他各种控制帧（例如一种控制帧是用来确定在环中是否还能接收新的站点）。

4）地址，校验和：与图 6-12 含义相同。

5）数据：没有长度限制。为了确保每个站点不至于太自私，使用一个令牌占用时间参数来限制发送时间。

6）FS（Frame Status：帧状态）：该域包含两个有意义的比特。当目标站点读完一帧之后，其中一个比特将被置位；当目标站点复制完一帧之后（如缓冲区未满），另一个比特被置位。源站点对这两个比特位进行解释：

00 = 目标站点没准备好或没开机；

01 = 目标站点准备好，但帧没被接受；

11 = 传输成功。

在令牌环中，总有一个帧在传输；如果没有数据传输，令牌将继续传输。站点一个比特接一个比特地阅读并以一个比特的时延向下一段传递。每个站点的作用就像一个重复器，涉及到的站点复制数据，经过一个循环发送端将数据从环中删除。

由于环一直在运行，令牌环的帧不需要像以太网那样为接收端的同步设置前导。

FDDI（Fiber Distributed Data Interface）与令牌环的关系就像快速以太网与以太网的关系。FDDI 的主要原理（在 ISO 9314 中标准化）与令牌环相同，但是 FDDI 使用 LWL 并且在传输距离大于 200km 时支持最多 500 站点，传输速率可达 100MBit/s。通常 FDDI 像 DQDB 一样作为 MAN 骨干网使用。FDDI 也支持具有实时要求的数据传输，如话音传输。

出于价格和安全的考虑，FDDI 在光发送上使用 LED。电缆间连接采用两个不同方向的环，其中一个作为另一个的保护。如果两个环在同一个位置中断成两个（如因为火灾的缘故），相邻两个站点的光入口与出口就连接起来。这样两个环就能构成一个更长的环，如图 6-14 所示。

3. DQDB（Distributed Queue Dual Bus）

像 FDDI 一样，DQDB 适合用于 MAN，同 LAN 相比，DQDB 不论是传输速率还是可以拼接起来的传输距离都有明显的提高。DQDB 可以以 140MBit/s 的速率服务

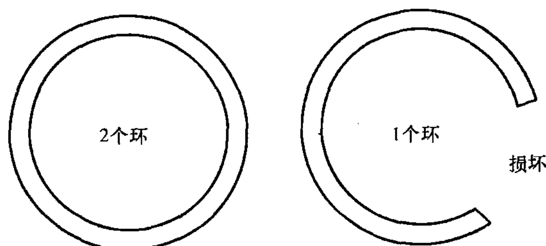


图6-14 一个FDDI双环的自愈

直径为50km的范围。

DQDB有两条平行的总线，所有的站点都连接在这两条总线上，如图6-15所示。每条总线的始端（Head）产生长度为53B的固定长度的信元，这些信元沿着总线向末端传送并在那里被移出。这样在两条总线上，信元沿着相反的方向传输。与LAN不同，DQDB使用固定长度的短信元，因此人们并不把它叫做帧。

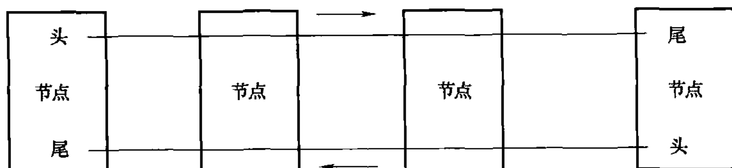


图6-15 MAN中DQDB的结构

如果一个站点想要传送数据，它必须首先知道目标站点在它的左边还是右边，这样它才能选择正确的总线并将一个空闲的信元放在上面。

这样做的问题是离总线的始端近的节点经常会包含许多空闲的信元，而这些信元却不能简单地被使用。通常对这些信元的使用采用相同权重的原则进行预留，它们不像可以实际用来传输的总线。每个站点为每个沿相反方向通过自己的被预留的信元增加计时器的数值，而为每个与想要传输的方向相同的空闲信元减少计数器的值，这样计数器就能给出不满足传输要求的下行流的数量。在另一个传输方向有另外一个计数器来做同样的处理。如果一个站点想要发送数据，它就在一条总线上预留一个信元，观察计数器的值，并在与传输方向一致的总线上让一定数量的空闲信元经过，然后占用下一个空闲信元。使用这种方法无需特别管制机制就可以实现一个公平的系统。

也有这样一种情况，在始端中会周期性地预留一个信元。因此，DQDB也适用于传送话音、视频等具有实时要求的业务。

如果总线的头和尾都放在同一个站点上，那么在物理上就形成一个双向环，但在逻辑上还是一个总线。目前每个站点都可以产生和接收信元，并且还能实现一个额外的能对环进行配置，即规定某些站点功能的算法。此外，目前当网络中出现意外站点时网络可以进行自愈，这样损坏站点的左、右两个站点都可以被配置成末端节点，如图6-16所示。

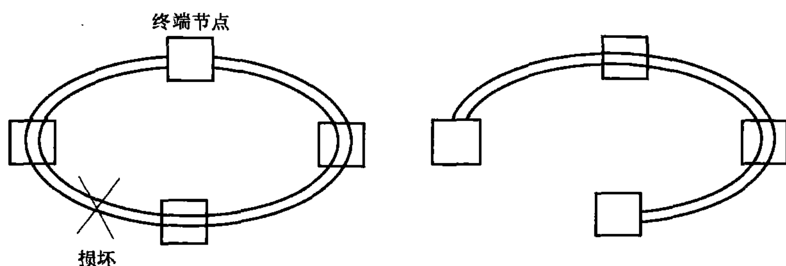


图 6-16 出现损坏时 DQDB 环的重配置

根据图 6-11，以太网、令牌环和 DQDB 都使用相同的 LLC 层的标准，因此它们可以通过桥连接起来。这也是 DQDB 的另一个主要的应用，即将远处的 LAN 连接起来，如图 6-17 所示。

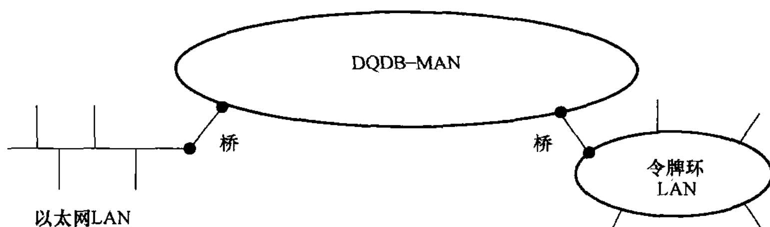


图 6-17 通过 MAN 将两个 LAN 连接起来

6.3 WAN 技术

6.3.1 交换

通常远距离网络指的是由国家企业垄断运营的、范围可达整个世界的电话网（美国例外）。这些网络主要用来进行话音通信，采用的技术（目前还是）电路交换（参见本节第 1 部分）和多路复用（参见 6.3.2 节）。与之相反的是，现在的广域数据网采用分组交换（参见本节第 2 部分），使用的是采用信元交换的 ATM 网络（参见 6.3.5 节）。当今 WAN 可以由私人企业运营。

交换指的是动态地分配网络的能力，根据分配方式的不同有不同的交换方式：

1）电路交换：用户在连接存在的整个时间内独占一条电路，如一条 64kBit/s 的电路，因此电路交换最适合实时业务。此时，双方对话时网络的利用率最低。

2）分组交换：信息被分成不同长度的分组，例如最大分组长度可为 64KB。在网络中每个分组都单独通过网络，以数据报或虚电路的方式被处理。分组交换最大的优点是电路利用率明显比电路交换高，但不适合实时业务，如话音。分组交换最大的应用是数据业务，最明显的例子就是互联网。

3）信元交换：信息被分成固定长度的、被称为信元的短分组。同分组交换不同，信元交换节点较复杂，价格较高，因此信元交换也适合实时业务，它综合了电路交换与分组交换的优点。ATM 采用信元交换。人们希望从长远上来说只需讨论一种统一的网络技术即可。

交换与多路复用位于 OSI 参考模型的第三层。

1. 电路交换

在电路交换情况下，首先要建立一个连接（由用户 A 初始化），然后该连接存在的整个时间内都只被用户对 A/B 所拥有，待通信结束后连接被拆除，占用的资源被释放。国际范围的电话网就采用这种原理。

可以这样理解电路交换，在 A 与 B 之间由多段中继构成一个电流环路。但事实上并不存在电流环路，因为在这之间信号已被数字化并被复用。用铁路网来比喻，电路交换的含义是在 A 与 B 之间所有各段铁路的交叉点都被正确地安放，并且所有的信号灯都是绿的。

电路交换的劣势很明显：尽管由于只有一列火车，多段铁路都是空闲的，但铁路网在同一时刻不能让多列火车同时通过。电路交换的优势是整个系统很简单，因此电路交换已被广泛地应用。

交叉点的位置（即交换机）及费率的确定需要信令，即在各交换机之间的信息流。ISDN 网络交换机之间使用 SS7（No. 7 信令系统）。SS7 信令使用 2MBit/s PCM 帧中被预留的第 16 时隙，如图 6-18 所示。在该时隙中同时传递一组用户信道的信令（common channel signaling，共路信令）。出于安全的考虑，每个交换机必须有两个以上、互不相关的信令信道可用。这些信令信道一起构成一个独立的、互联的、层次化的信令网。

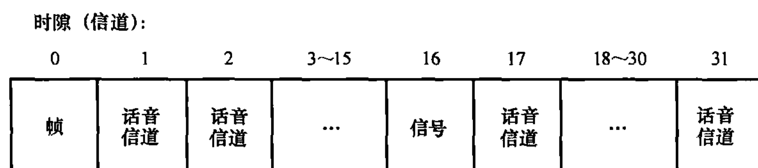


图 6-18 PCM30 路系统的帧结构

在电话网的初始阶段，交换机的作用只像个转换插座，接线员将希望的连接插到相关的插座中。此时信令信息通过用户 A 口头通知给接线员。

出于经济和数据保护的原因，信令系统被自动化。Almon Strowger，美国的一个葬礼企业主曾经很愤怒他的许多订单都落到了一个竞争对手手里，仅仅因为这个竞争对手的夫人是当地的电话接线员。为此他发明了第一台旋转式交换机，以便能合理地赶走这个竞争对手的夫人。此后由惰性气体保护的中继器接管了这项工作，在 100 年中它始终是直流电路连接。当交换机变成了是受程序控制之后，仍然是这样。

当今传输和交换都已数字化。所有的数据流都按时间进行复用，目的是共同

使用30个64kBit/s的信道。每帧包含30个8Bit的话音信道，另外还有8Bit用于帧同步，8Bit作为信令。因此每帧包含32B=256Bit。因为每秒中必须传递8000帧（话音采样频率），因此总的数据传输速率为2.048MBit/s。图6-18所示为这些帧的组成。这个系统被称为PCM30路系统，并在欧洲使用。在美国和日本使用24路64kBit/s的系统，每帧的速率为1.544MBit/s。

通过交换机连接入口端TDM线路和出口端TDM线路，如图6-19所示。

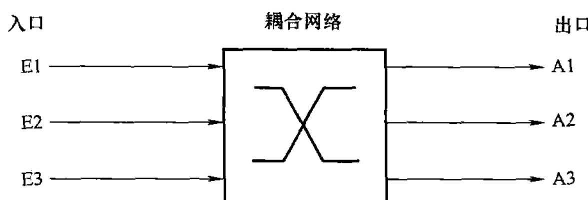


图6-19 交换机连接TDM流（所示仅为单方向的连接）

交换机的中心是一个网络耦合器，它需完成两个任务：

1）空间上的耦合：入口电路时隙到出口电路的转换。

2）时间上的耦合：同一帧内不同时隙间的转换。例如图6-19中入口E1的第12时隙及入口E2的第12时隙都需映射到出口A3继续转发，因此其中的一个时隙分配必须进行转换。

如果大于30个信道都需由一个出口中继转发，那么交换机就会出现过载。这时用户A就会收听到忙音，不论B是否正在用电话。如果这种情况经常出现，网络就必须进行扩容，在出口端配置另外一条TDM电路。当然这两条电路也需根据交换机的情况进行复用。

如果出口电路没有过载，但从入口到出口的映射却不能成功，此时交换机被锁死。采用适当的技术可以实现无锁死的网络耦合。

电话网是分层次的，相关的交换机分别称为用户交换机、节点交换机、中继交换机、国际交换机，但所有交换机的工作原理都是相同的。

集线器是能力有限的交换机，它需由节点或中继交换机进行远程控制。

电路交换的一个最大的缺点是它的效率很低，即它没能充分利用传输的能力。在一个电话通话过程中，从A到B和从B到A两个信道都被一直预留。而这两个信道的利用率都大约只有40%，因为从头到尾只有一个人在讲话，而且其中还包括一些停顿。在4.1.3节的结尾部分介绍了用DSI方法及TASI方法可以对此进行的一些改进。在互联网会话时，情况更遭，因为绝大多数情况下，由用户到网络方向的信道是空闲的。从这种意义上看，分组交换的效率较高。

在电路交换中，信道占用及非用户信息交换都需要被计费，这从运营的角度上来说是很合理的，因为运营商并不能将已被占用的信道用作它用。

2. 分组交换（数据报、虚电路）

像E-mail这样没有实时要求的业务使用的是分组交换。此时消息通过传输层

(OSI参考模型的第4层)被分成多个分组(通过使用分片、分段技术),然后从一个节点到另一个节点在网络中传输。只要下一段有空闲的传输能力,分组就可以被传输,而不需要像电路交换那样从发端到收端的每段线路都得空闲。

每个中间节点都必须存储分组,直到下一段网络有空闲资源(即所谓的存储、转发)。因为此时时延与网络的负载相关,因此如果不采取其他的措施,对具有实时要求的应用是不适合的。为此采用一种时间上的负载均衡技术(流量整形)。

数据源每秒可能发送许多分组,也可能只发几个。因此网络在发送分组少的应用分组之间可以传递其他业务的分组。分组交换与电路交换相比有许多优点:

1)按照用户实际发送的有效数据进行计费,而不是按照时间。

2)网络的传输能力能够被充分地利用,因此价格较便宜。

3)对用户来说信道是可扩展的,即可以设置成多种中间速率。

4)每个终端设备可以按照自己的数据传输速率工作,也可以将网络中不同的段设置成具有不同的速率。

分组交换分成两类:

1)无连接方式:即没有连接的建立过程→数据报。

2)有连接方式:即在发送第一个分组之前建立连接,发送完最后一个分组之后将连接拆除→虚电路。

邮政是一个无连接的业务,相反,以电路交换为基础的电话业务是面向连接的业务。

数据报的每个分组都包含发送者及目的地址,并从一个节点向另一个节点传递。每个节点读取目的地址,并使用最优的路由,将该分组发向有空闲资源的下一跳。最优路由发生变化的情况很有可能发生,因此后续的分组可能会走另外一条路,甚至可能会比前面的分组先到达目的地。OSI协议栈的第4层负责检查是否所有的分组都已到达目的地,并按正确的顺序整理出来。因此在第4层的分组头中要将分组进行编号。至于哪条路由最优,这是路由过程的结果(选路,参见6.3.3节)。

而虚电路则同电路交换类似,首先建立一个连接。但此时并不是真正建立一条信道,而仅仅是在每个节点确定了分组的去向。实际上在每个节点存储一个表格,阅读表格就知道哪个分组发送到哪个目的地。因此,一个连接的所有分组就能被动地采用相同的路由,按照正确的顺序到达目的地。

打个比方来说,可以用一列火车的车厢来理解分组交换。这列火车可以理解成消息,而每节车厢可以想像成一个分组。在数据报的情况下,每节车厢有一名司机,司机将车厢开出,到达每个路口时根据站在那的交通警察的指挥,将车厢开往正确的目的地。交通警察之间使用对讲机相互联系,知道哪儿交通拥挤,因此可以指挥司机走一个最佳的路线到达目的地。

在虚电路的情况下,事先发一个声明出去,该声明达到每个路口之后在该路口设置一个路标,此时也需要交警的配合。最后列车的司机收到声明,并按路标将车开到目的地。

同数据报相比，虚电路有如下的优点：

1）分组经历的开销较小，因为在每个节点路由已知。开销仅仅是连接号，而不是目的地址和序列号。

2）节点的负担较小，因为它不需要为每个分组都选择最优路由。

3）第 4 层的负担也变得较小。

数据报的最大优点是它具有绕过突发网络瓶颈的能力；而虚连接却不具备这种能力，可能只因为网络中的某一点拥塞。

ATM（异步传输模式，参见 6.3.5 节）被认为是宽带网络的代表。分组网络要求效率，应用要求实时性。虚电路能解决这个问题，并且节点和第 4 层还能变得简单和快速。所有的分组可以设置成具有相同的长度（ATM 是 53B），称为信元。这样节点可以进一步地精简，并实现话音和数据的有效的实时传输。局域范围的网络负载过大可以通过预留的方式解决，这也需要建立虚连接。因为业务有平均、最大的传输速率需求及可忍受的最大传输时延，网络可以预估它的负载，并可以拒绝或迂回连接需求。绝大多数情况下可以避免网络的过载。

分组可以根据它的目的地址及连接所属的对象进行标识，例如一个用户可以同时浏览网络或发邮件。此时他物理上使用一个信道，但逻辑上却是两个。

现在还得解释一下可靠业务和非可靠业务这两个说法。非可靠业务的分组可能出现丢失、篡改或复制的现象。这样高层必须负责在非可靠的低层的服务的基础上为用户提供可靠的业务。而可靠业务指的是由网络层负责处理分组的丢失、复制及篡改。

面向连接、无连接、可靠及非可靠这些特性可以有 4 种组合，其中最主要的是面向连接的可靠业务及无连接的非可靠业务。

图 6-20 所示为 IP 协议栈（IP，参考模型指的是图 6-7）数据报分组的构成方式的一个例子，从这个例子可以看出上述数据报的原理是如何实现的。

分组头由 20B 的固定长度部分（图 6-20 中的前 5 行）及长度可变的可选部分组成。它们的含义分别是：

1）版本：IP 的版本。这样允许快速引进新的协议版本。

2）IHL：用 32Bit 为单位表示的分组的长度。最小值为 5（即没有可选部分），最大值为 15。

3）TOS（Type Of Service，业务类型）：由数据的内容（业务）来确定，如话音业务（速度比准确性更重要），或者数据业务（准确性比速度重要）等。使用这个域在网络出现拥塞时可以作出适当的妥协。

4）总长度：分组的总长度（包括分组头和分组体），单位是 B。一个分组的最大长度为 65535B。

5）标识：用来表明属于同一个消息，被分成多个的分组。

6）UDM：U 表示未用，D 表示未分段（网络不允许分段，例如因为目标不在一起），M 表示还有其他分段（除最后一个分段外的所有分段都设置该位）。

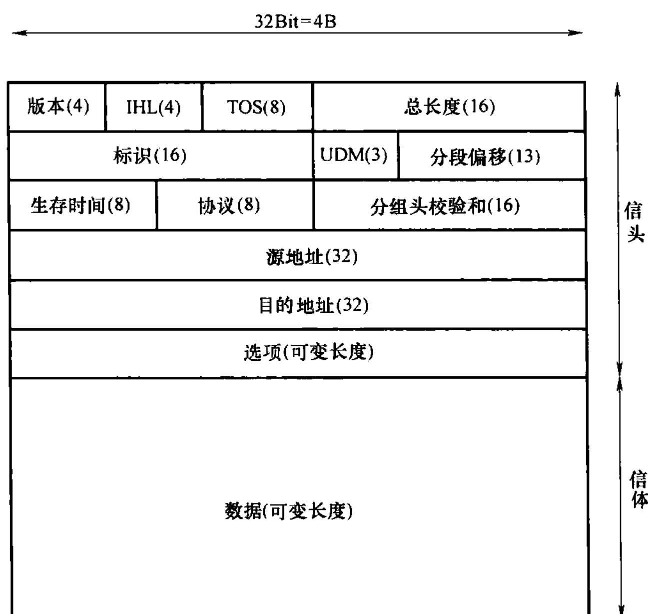


图 6-20 IPv4 分组的构成（版本 4，括号内为每个域的比特数）

7）分段偏移：设置每个分段的偏移量。它的目的是在接收端可以按照正确的顺序将各个分段重组。

8）生存时间：限制分组生存时间的计数器。经过每个节点时该值被减 1，到达 0 时分组被丢弃，并向发送端发送一个告警。设置该域的目的是防止分组由于错误路由而一直在网络中传递。

9）协议：用以表明在接收端网络层如何开始处理这些分组。

10）分组头校验和：对分组头进行校验。每个节点必须重新计算该域的值，因为生存时间域的值在不断地发生变化。

11）源地址：发送者的地址。

12）目的地址：目的地的地址。

13）选项：为其他服务、服务注册及未来版本预留。

多年来第 4 版本的 IP 在互联网上被广泛地应用，由于它的成功，它已经达到了使用的极限。最大的问题是 32Bit 的地址长度只有 2^{32} 种可能性。在发展的初始阶段，人们只是希望它能尽量满足所有的要求，但是今天它却不能覆盖整个地球。为了能简化路由，将网络分层进行组织，这样能使更多的终端用户（host）组织成不同的子网。因此 IP 能够区分子网地址和终端地址，通过 5 类（由 A ~ E 来标识）地址来标识（许多网络使用较短的终端地址，或相反终端地址长度较长）。这些地址在世界范围内是独立的，并由一个中央机构来分配。由于被分配了过多的组地址，并且这些组地址并没有充分地利用，这使得必须分配新的地址。因为这些组

地址漏洞并不能简单地被陌生的终端使用，否则就破坏了地址的分组组织结构，因此出现了第 6 版本的 IP。图 6-21 给出了 IPv6 分组的构成，其中各个域的含义如下：

1）版本：同 IPv4 该域的含义相同，此时该域的值为 6。

2）优先级：同 IPv4 中 TOS 域的值含义相同。

3）流标识：流指的是具有相同的源的所有的分组的集合（指应用进程，并不是指机器）。流标识可以识别一个节点（这里指路由器）所处理的同类分组，因此适用于实时业务。

4）负载长度：有用数据的长度，单位是 B。在 IPv4 情况下，总长度域指的是整个分组的长度，包括分组头。对 IPv6 来说，分组头具有固定的长度，因此 IHL 域被省略掉。

5）下一个头：相当于 IPv4 域的协议域。由于 IPv6 没有选项域，代替它的是所谓的扩展头。扩展头可用下一个头域来连接起来。同 IPv4 相比，使用扩展头还可以实现许多安全相关的功能。

6）跳数限制：相当于 IPv4 的生存时间域。

7）源地址：发送者地址。

8）目的地址：接收端的地址。

IPv6 的分组头长度比 IPv4 20B 分组头长度多 40B，但 IPv6 域的数量却相对较少，这样可以减少路由器对分组的处理时间。

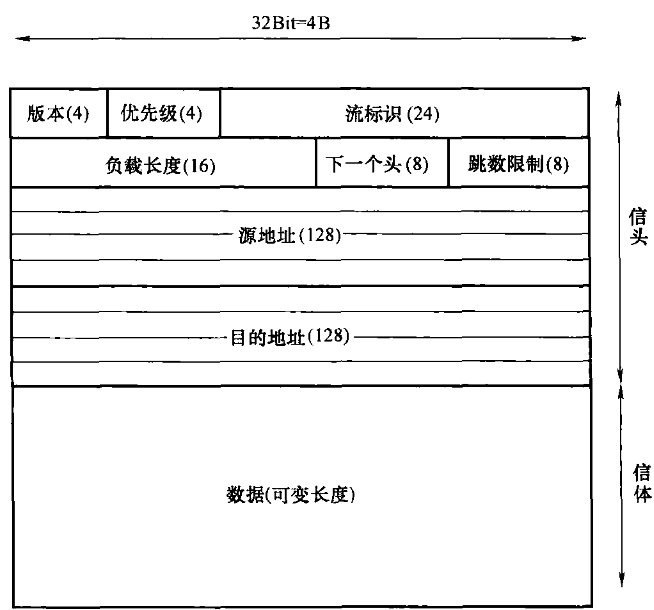


图 6-21 IPv6 分组的构成（版本 6，括号内为每个域的比特数）

IPv6 的形成曾经经常遭到激烈的质疑。引入 IPv6 的主要目的是扩大地址空间，除此之外还包括简化分组头以减少路由器的处理时间，因此还去掉了分组头中的校验和部分，因为目前采用光线来传输数据，物理层的误码率很低，端到端的差错完全可以通过传输层协议来处理。

这里还得说一下使用跳数来限制域的规定。IPv6 同 IPv4 相同，最多可以允许 255 跳，这却与巨大的地址空间相矛盾。流行的观点是将来随着远距离连接的增多，跳数的数目并不增加。

IP 目前主要应用在 LAN 范围，例如通过以太网的帧来进行传输。

6.3.2 复用

如果满足下面的前提，复用是减少费用的一个很有效的方法：

1）多个逻辑连接走相同的物理路线。在实际上仅能实现部分复用，通常在带宽比较高的信道上仅是进行粗粒度的划分，而在带宽比较窄的地方进行细粒度的划分。

2）信道的传输能力还没达到理论上的极限。

3）在信道上的花费与信道的传输速率不成比例。因此线路的捆绑当然也符合这个原理：在同一个房间里布置两条同轴电缆的费用肯定比在两个房间里各自布置一条的费用低（可以看成房间复用）。布置 1GBit/s 的 LWL 线路的费用一定比布置两条 500MBit/s 的光纤的费用低（可以看成是时间复用）。

从技术来说，复用包括频分复用（FDM）、时分复用（TDM）和空分复用。

1. 载频网

有线 FDM 系统也被称为载频网。以前它被用在电话网中，现在几乎已经没有了，因此对它的描述比较简短。但在以前，它的技术简直是个巧妙之举。

这个系统基于模拟技术，传输媒体采用同轴电缆，因为那时还没有 LWL。高频电子技术那时还比较昂贵，因此人们使用节省带宽的，但对话音已经足够了的 SSB 调制。因为有干扰很小的同轴电缆，因此可以不需要调制增益。

为了进一步节省电子电路的开销，并没有对每个话音信道都进行调制，而是采用了分层化的方式：

1）首先将 12 个话音信道在 60～108kHz 范围内分成主要组。这需要很陡的滤波器，给每个信道分配 3.1kHz 的带宽。

2）频率范围为 312～552kHz 的 5 个主要组形成有 60 个信道的次要组。

3）15 个次要组构成有 900 个信道的基本主群组。

4）3 个基本主群组共有 2700 个信道，总带宽为 12MHz。

5）最后 4 个基本主群组共有 10800 个信道，总带宽为 60MHz，但放大器的间距只有 1.5km。

2. 准同步数字序列（PDH）

PDH 网络使用时分复用（TDM）方法，这从名字中的“数字”就已经能够看

出来一些。集群，即由多个子数据流构成的数据流，天生就有很高的数据传输速率。相应的，人们使用同步传输，参见表2-1。

准同步意味着“非常，但还没有完全同步”，这个问题可以用图1-49来解释。用户A1和A2的信号一起集中到了K2-K3这一段落上。遗憾的是从A1出来的信号在交换机AZ1进行数字化，而A2出来的信号在AZ3进行数字化。每个交换机有它自己的时钟振荡器。因为归一化的频率到处都有，有效频率之间会有一定的偏差。如果现在将数据流以一个固定的模式混在一起，那么偶尔会产生所谓的比特滑动（比特丢失）。在2.2节中的异步传输中我们会遇到相同的问题。

一种滑动离合器能减轻这个问题，在TDM系统中将它称为填充。遗憾的是这个概念在透明的同步传输中已经被使用了（2.2节），但这里却有不同的含义。

如果将4个数据流混合在一起，新集群的时钟频率会比原始流的时钟频率的4倍还要多一些。因此上一层帧会包含多余的比特，即所谓的填充比特。需要时可用到它，否则的话就作为无意义的填充比特一起来传输。

这个问题在复用不同的2MBit/s的子数据流时才会出现。如果这些2MBit/s的子数据流来源于同一个交换机，那么不会产生比特滑动。欧洲的PDH系统使用层次结构，它们都基于4个子流，参见表6-3。

不同层次间的过渡就是复用和解复用。有这样的复用器，可以将16个E1组合成一个E3。人们目前很喜欢将E2跳过，因为与E3相比，它在价格上没有什么优势。E5已经被SDH技术（参见下节）代替。在表6-3中E1、E3、E4还有较大的意义。

美国和日本使用与表6-3不同的层次结构，图6-22给出了3种所使用的PDH系统的构成。

表6-3 ITU-T给出的准同步数字序列（PDH）的特性

ITU-T的描述	64kBit/s话音信道的数量	精确数据传输速率/(MBit/s)	简化的数据传输速率/(MBit/s)
E5	7680	565.148	565
E4	1920	139.264	140
E3	480	34.368	34
E2	120	8.448	8
E1	30	2.048	2

PDH技术的最大缺点就是填充。为此单个比特的时间位置是可变的，因为每一层根据自己的需要都会有填充。因此，在一帧中的数字1234可能不会一直属于一个信道。如果想要重新构造信道（例如在图1-49中的节点K2和K3中），那么需要将整个数据流都解复用成希望的数据传输速率，然后重新构造，然后再进行复用。图6-23描述了从E4中分离出E1所需要的巨大的复杂性。这种Add Drop Multiplexing（ADM，分插复用）功能可以用在例如将2MBit/s的租用线给一个大公司自己的WAN使用时

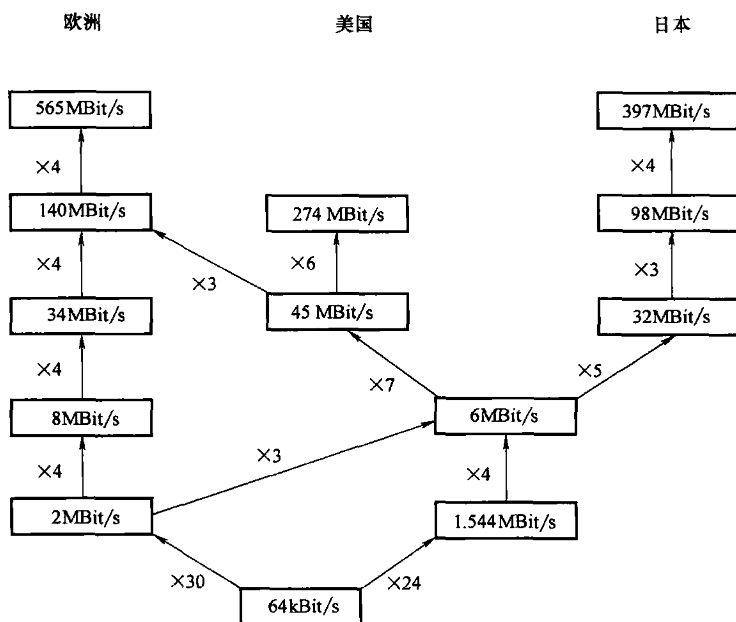


图 6-22　传统 PDH 系统的构成

的情况。填充的缺点以及对世界范围内一个统一的标准的希望，一起促进了 SDH 技术的发展。

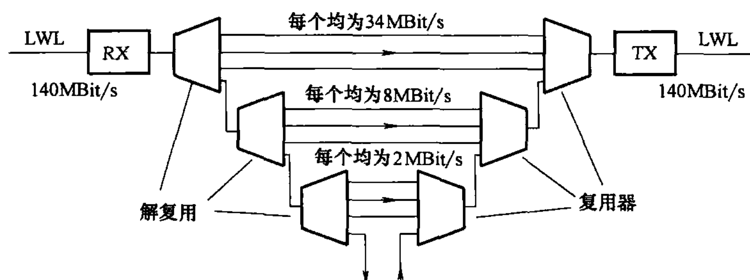


图 6-23　PDH 系统中的加减多路复用器（只描述了一个方向）

在一个大交换机中，不仅对图 6-23 中所示的单个集群进行解复用，还将整个集群都解复用到 E1 上。然后这些 2MBit/s 的信号被送到交换机中，接着又被进行复用。租用线路（标准线路）会绕过交换机。这里涉及的不是租用的“线路”，而只是没有被交换的 TDM 束的时隙。

3. 同步数字序列（SDH）

1990 年 SDH 技术在 ITU-T 建议 G. 707 到 G. 709 中被确定。美国人将这种技术称为 SONET（Synchronous Optical Network），这指明了它主要用 LWL 来进行传输。但除此之外，也通过定向无线线路来传输。由于其灵活性和有效性，SDH 网络是未来具

有较高复用层次的传输网络。对于 140MBit/s 以下的低速率，PDH 还将会使用很长的时间。

相比于 PDH，SDH 有很多的优点：

1）世界范围内的标准使跨边界的流量不使用接口转换器成为可能。

2）使用差分 ADM 使简单的信道重构（提取出或增加单个的 SDH 和 PDH 束）成为可能。

3）新的线路电路板无需像在 PDH 中那样需要大范围的复杂的焊接工作，而是使用软件控制即可完成，即所谓的交叉连接。

4）网络监测和控制所需的额外的传输能力已经被考虑在内。

5）一方面，SDH 可以传输老的 PDH 帧（透明传输）；另一方面，由于较快的重构能力，也为未来的在 WAN/MAN 范围内的使用分组交换的 ATM 网络打好了基础。

SDH 网络需要一个中央时钟提供者。这可以在全国的范围内实现，可以用铯原子钟作为统一的频率标准。

最后，SDH 当然也基于帧的传输。但在 SDH 的术语中，帧被称为 STM（Synchronous Transport Modul）。

1）模式 STM-1（Transportmodul Level 1）每秒钟被传输 8000 次。它由 9 行，每行 270B，每个字节 8Bit 组成。因此产生的速率为 155.52MBit/s。

2）4 个 STM-1 复用成一个速率为 622.08MBit/s 的 STM-4。

3）4 个 STM-4 复用成一个速率为 2488.32MBit/s 的 STM-16。

每个传输模式由多个容器和一个头开销组成。一个 STM-1 的容器可以封装任何数据，包括图 6-22 中的 PDH 群。

在头开销中，用一个指针解释清楚该模式的构成，指针被 ADM 来使用。在 SDH 中帧是按结构化的方式组织的，与 PDH 的面向数据位的构成方式不同。在 SDH 中也有填充，但它比 PDH 巧妙，是用指针来组织的，并因此允许单独地对每个数据流进行重构。

SDH 为 PDH 或其他速率到 140MBit/s 的宽带源提供透明的信道。通常情况下网络是由层次化的（双）环构成，考虑到安全因素，每个环都用多于两个的连接与上层环相连（冗余），如图 6-24 所示。在每一层上都有网络管理。以前人们使用星形或树状的拓扑结构，后来都连在了一起。目前人们喜欢使用环形结构。

SDH 技术的核心单元是 ADM 和交叉连接，软件控制使灵活的网络重新配置称为可能，相反 PDH 技术还受静态的硬件技术的年龄的限制。

ADM 有两个光端口和馈线，通过两个光端口可以将 ADM 组成如图 6-24 所示的环形。馈线是 E1、E3 或 E4 或低层的 STM。交叉连接是与交换机功能相似的无阻塞的耦合。但它们不给单个信道提供交换业务（它由与终端用户无关的信令来控制），而给束（E1、E3、E4 或 STM）提供交换。它由网管中心来控制，当网络出现过载时，允许快速的、软件控制的重配置和网络适配。图 6-25 所示为一种网络配置情况。

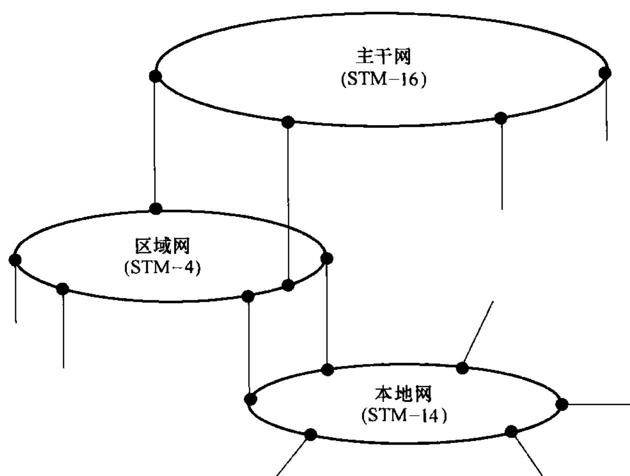


图 6-24　SDH 网络的结构

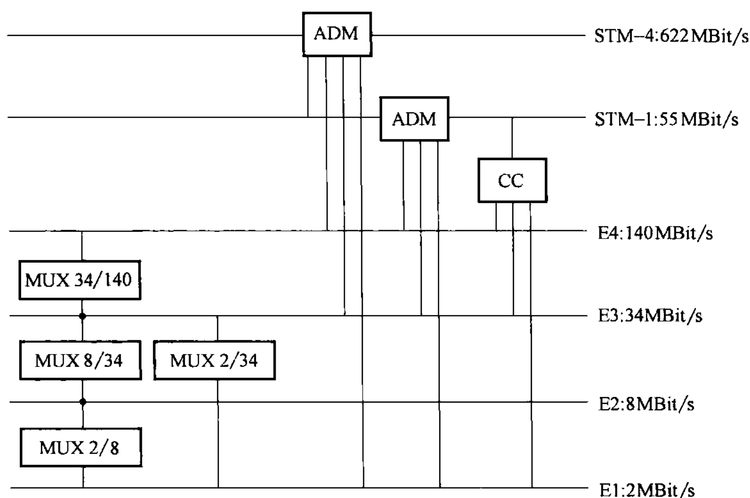


图 6-25　PDH 与 SDH 的混用

MUX—多路复用器　ADM—加减多路复用器　CC—交叉连接

6.3.3　路由

路由意味着在网络中或与网络的连接中进行路径的选择。静态交换网，如传统的电话网中的路由是很简单的，使用号码就能清楚地确定在整个网络中的路径。在传输路径中的交换机不断地翻译部分被叫号码。例如一个 TVA（企业交换机，家庭小交换机）能够根据所拨号码首位的 0，得知用户希望接的是外线，然后将线路接往公网。相连的公网的交换机继续分析号码，如寻找下一位的 0，并将线路接往下一级网络。

路由在像现代分组数据网这样的动态网络中是不太容易的。为了将分组发往正确的目的地，路由器必须知道现在连接的是谁（主机的地址）、在哪（路由器的地址）。在数据报的情况下必须为每个分组选择路由；在虚电路的情况下为每个连接（会话路由）建立路由，之后这个会话的所有分组都走相同的路由。

电路交换、虚电路、数据报也可以根据选择路径的时间点进行区分：电路交换路径的选择是在网络建立时进行的；虚电路是在连接建立过程中；而数据报针对每个分组临时建立路由。因此在静态交换网络中讨论数据报意义不大。

路由实际上是求解一个优化问题，根据不同的目的有不同的路由算法。如优化的目标可以是最大化网络的吞吐量，也可以是最小化分组的时延或是最小化分组经过的跳数。

在具体实现中，每个路由器都有一个路由表，记录如何到达每个主机，因此问题就变成如何维护和管理这个表格。在实际实现时，这涉及到路由器的数据库与实时操作系统。

路由表的自动更新可以通过例如每台主机在网络中的自我介绍来实现。也就是说，主机向它的下一个路由器发送一个分组；路由器将该主机的地址填到路由表中，将自己的地址填入到分组中，并将该分组发往与之相连的所有路由器，除了接收到该分组的主机。所有后续的路由器均重复同样的动作。由此可以发现多个可能的路由，根据优化算法可以选出最优的一条路由。在复杂的网络中可能会出现个别分组在网络中出现死循环的现象，这些分组必须被识别出并且从网络中移除。

在大的网络中，路由表相应也会很大，因此相应的查找时间也较长。将网络按层次化的方式分割成子网有助于改进这种现象。

由于网络的状态和拓扑结构在不断地变化（主机在不断地开关，链路偶尔会断掉，网络会扩容），路由信息隔一定时间就得更新一次（数量级从秒到分）。根据第 3 层的需求，第 1 层和第 2 层会有一定的开销。

关于路由的详细内容可以参见相关的专业文献。

6.3.4 流量控制

流量控制（flow control）关心的是一个高速发送端的数据不致于将较慢的接收端淹没。在数字化的电话网中（电路交换）不存在这种问题，但在分组交换的数据网中却存在这种问题。

流量控制意味着接收端能够控制发送端的数据输出速率，因此需要一个反向的反馈信息。第 2 层使用 ARQ 纠错的退出过程就与流量控制有关。端到端的纠错这种方法是不可行的，因为到发送端获悉数据发送得太快时，已经有许多数据分组发送出去，接收端仍然会被淹没。

流量控制虽然可以端到端地进行，但由于时延太长，不是很有效。较好的办法是让中间节点缓冲数据，进行点到点的流量控制。这样做还有另一个优点，因为即使真正的接收端还有足够的接收能力，网络中的节点本身也可能出现过载。例如，如果所有三个入口线路都得将数据分组交换到同一个出口线路，可能就会出现这种现象。

在实际中经常使用的方法是接收端时常返回一个反馈信息，告诉发送端可以发送多少个分组。

6.3.5 通信系统示例

1. 帧中继

帧中继是一个简化的、使用虚电路来加快分组交换的系统，通常使用 X.25 协议，主要具有如下的特性：

1）负责连接建立和拆除的分组（未编号的 HDLC 帧，如图 6-9 所示）同数据分组使用相同的信道（带内信令）。

2）流量控制由第 2 层和第 3 层来完成。

第 2 层的差错安全意味着，每段都单独实现纠错。这意味着网络中的节点需要为此进行额外的工作，从而导致有用数据传输的时延。第 3 层和第 4 层的数据传输速率由于差错处理并没有增加。

这些额外的工作对误码率较高的信道是比较适合的，也正是因为这个原因才引入的这种方法。与此相反，LWL 连接的误码率很低，因此在第 2 层采用纠错措施的缺点大于优点。因此帧中继没有流量控制。可以根据需要在第 4 层采用端到端的流量控制。有效数据的传输速率可达 2MBit/s。

2. ATM

ATM 是一个统一的宽带网络。在 6.3.1 节中已经介绍了为什么需要引入使用被称为信元的固定长度分组、基于虚电路的分组交换。对数据传输来说，信元越长，信元头等开销所占的比例越小，因此越有利。但对像话音、可视电话、视频会议等实时应用来说，最好使用较短的信元。在对 ATM 进行标准化时，美国人推举使用 64B 的有效载荷，因为他们希望使用 ATM 进行数据传输。相反欧洲人则希望使用 ATM 进行低成本的话音传输。最后进行了折中，规定有效负载为 48B。加上 5B 的信元头开销，ATM 的信元长度为 53B，如图 6-26 所示。

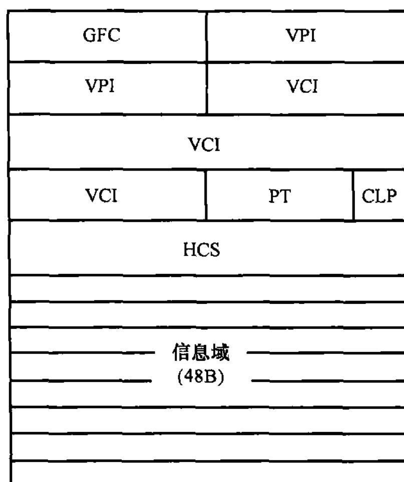


图 6-26 ATM 信元的构成，每行表示一个字节

图 6-26 中个域的含义如下：

1）GFC（Generic Flow Control，通用流量控制）：仅为主机和网络之间的连接而考虑，很少使用。

2）VPI（Virtual Path Identifier，虚通路标识）：因为分组的交叉存取而使传输介质（例如 LWL）产生了多个逻辑连接。这些逻辑连接被称为虚通路，也就相当于虚电路中的路标。

3）VCI（Virtual Channel Identifier，虚信道标识）：每个虚通路可以被分成多个具有相同网络路径的虚信道。这在例如立体声广播传输中很有用——可以事先预留一条具有两个信

道的路径。网络中的路由器只负责处理这条路径（VPI）即可，两条信道（VCI）由终端负责处理即可。

4）PT（Payload Type，负载类型）：根据传输的内容（由用户指定）和传输的难度（由网络指定）来确定。

5）CLP（Cell Loss Priority，信元丢失优先级）：信元的优先级。出现问题时 ATM 交换机首先丢弃低优先级的信元。

6）HCS（Header Check Sequence，信头校验序列）：信元头校验和。

可以将 ATM 系统想像成一个一直在运行并且可以根据需要来装卸货物的缆车系统。如果没有准备好有用数据，就传输空信元。话音数据的传送只是占用其中的一些信元，但要求这些信元之间有固定的间隔。相反，压缩视频的传输占用间隔不等的多个信元。较大的分组需被分段，这样它们才能用信元来传输。ATM 的优点如下：

1）可扩展性和灵活性：每个连接的信元数都是可变的，因此 ATM 对所有的业务都是适合的。

2）有效性：分组交换使线路的传输能力被充分地利用，此时不需用时间来计费，而是根据所传输的位数和占用的虚通路数来计费。

3）不依赖网络和应用。

ATM 的工作速率为 155.52MBit/s，正好适合 SDH 的 STM-1。此外高速的分组交换要求有高速的转接车站，也就是说 SDH 要与特殊的 ATM 交换机（路由器）相结合。因为 ATM 使用虚信道，ATM 信元的顺序由传输来确定，但虚连接则需要有一个建立的过程。在这过程中根据表 6-4 选择希望的业务类型，这些业务类型根据网络负载进行优化。此外，峰值信元速率、信元抖动等还需确定。超出连接建立过程中所承诺的传输能力的数据将被拒绝。网络则相应地检测用户的数据是否同预定的一致、是否超出了范围（policing）。

表 6-4　ATM 的业务类型

分类	说　明	应用举例
CBR	固定比特速率	话音
RT-VBR	可变比特率，实时	视频会议
NRT-VBR	可变比特率，非实时	视频点播
ABR	可用比特率	Internet 浏览
UBR	未指定比特率	数据传输

为了实现传输层的业务，需要根据应用指定一个适配器，这些被称为 AAL（ATM Adaptation Layer）。向上 AAL 的任务是保证所提供的业务已经准备好并隐藏 ATM 的具体机制。

根据实时/非实时、固定/可变传输速率、面向连接/面向非连接等判据可以将业务分类，有 8 种组合。其中 4 类已被支持，其他 4 类显得不是很重要。相应地有 4 种类型

的 AAL，见表 6-5。因为之前曾经有多种 AAL，因此表 6-5 所示的 AAL 逻辑并不是很清晰，如 TCP/IP 分组使用 AAL5 来封装到 ATM 信元中。

表 6-5　AAL 的业务类型

AAL-Typ	AAL1	AAL2	AAL3/4	AAL5
时实性	必要		不必要	
比特率	固定	可变		
面向连接	是		是/否	是

ATM 对丢失或被篡改的信元没有改正措施，也不包含任何流量控制功能。因为这两者对低成本、快速的 ATM 交换（主要用于实时应用）都有影响。如果有需要的话，这两个功能可由传输层来实现。

ATM 有它自己的参考模型，它比 OSI 模型更适合 ATM 所特有的环境。

由于特有的灵活性，ATM 已被广泛地应用，特别是在需要高传输速率的环境中。ATM 可以模拟 LAN，也可以用来连接不同的 LAN。ATM 最适合运行在 SDH 之上（不仅仅是因为 SDH 内在的管理功能），也可以运行在 PDH 之上。因此目前对 ATM 有很大的希望和期待，并且已有许多投资。

ATM 同以前的 DQDB 有相同的信元长度。两者都使用分组交换来传递多媒体信息，但由于使用预留技术，ATM 在网络负载过大时比 DQDB 表现得更好。

以太网和 FDDI/令牌环的 80% 都属于第 2 层技术。LAN 之间长距离的耦合可以通过私有网络的帮助来实现。这包括租用不同承载商的线路（以前是国家的 PTT，现在是私人企业）。出于对用户安全的考虑，也可以由电信部门铺设一条 $n \times 64$kBit/s 的线路。一个应用例如可以是发送一个文本文件。文本可以是在 Macintosh 机器上使用 Microsoft Word 产生，并通过以太网连接到网络上，从欧洲发出；而接收端却是在美国通过令牌环上网的 Windows 计算机，并且可以使用 Word Perfect 来处理接收到的文本。图 6-27 所示的是目前所采用的技术。

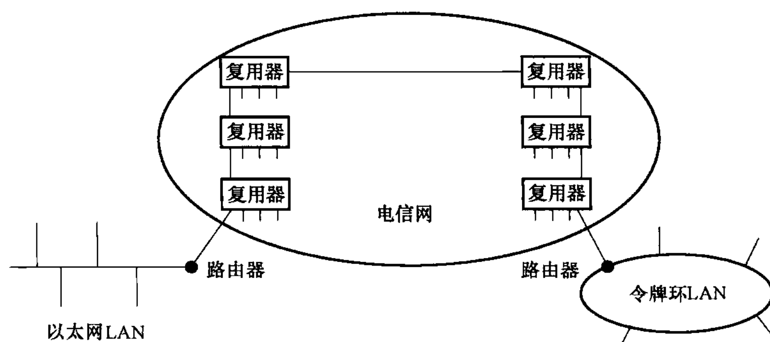


图 6-27　两个 LAN 通过公共网络（PDH 或 SDH）来连接

一个国际化的大企业（例如一个大银行）通信费用中的很大一部分都是付给了租用网络，而随着多媒体应用的发展，这个租用网络的移动性、容量、可扩展性等方面的能力却越来越不足。ATM 却潜在地可以帮助解决这些问题，并且不论是 WAN 还是 LAN 都可以采用这种技术。但 ATM 价格还很高，因此扩展的帧中继/吉比特以太网对 ATM 是个挑战。

6.4 协议与结构

6.4.1 TCP/IP

由 6.1 节的叙述我们知道，通信由不同层的协作来完成，其中的每一层都可由软件来实现，由此可以看出所有的软件都参与了数据通信。作为一个例子我们来看一下如图 6-7 所示的 TCP/IP 协议族。大致上可以将该协议族分成传输层协议（TCP）和互联网协议（IP）。图 6-28 所示为用户数据在不同层上的位置，该图可以与图 6-6、图 6-20及图 6-21 结合阅读。

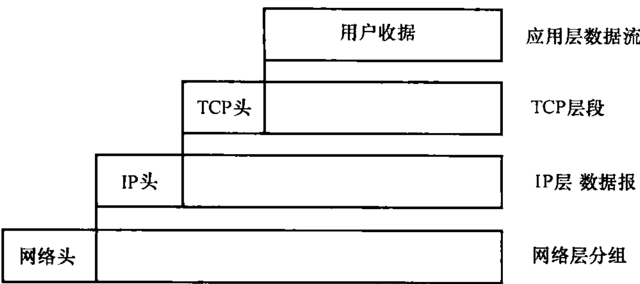


图 6-28 TCP/IP 结构中的数据单元

图 6-29 所示为部分协议及它们在 TCP/IP 结构中的位置。

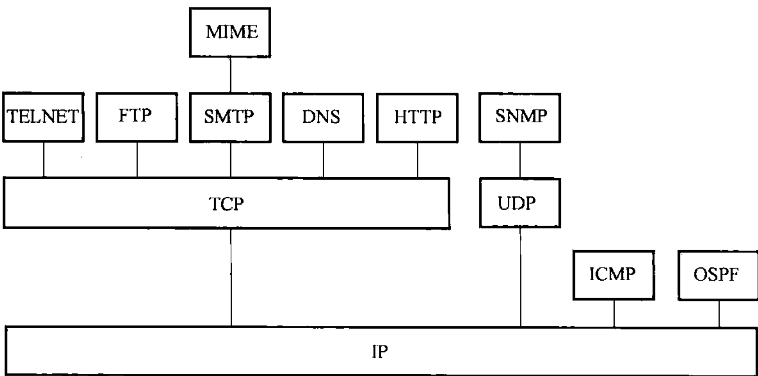


图 6-29 TCP/IP 协议族中的部分协议

下面是图6-29中所示协议的简要介绍，它们分成以下几组：

（1）应用层协议

1）TELNET 允许到远程计算机的远程登录。

2）FTP，File Transfer Protocol，文件传输协议。

3）SMTP，Simple Mail Transfer Protocol，简单邮件传输协议（E-mail）。

（2）IP层协议

1）IP，Internet Protocol。

2）OSPF，Open Shortest Path First，开放式最短路径优先。它是一个内部路由协议，负责根据事先定义好的准则（价格、时延、跳数等）来选择路由。

3）ICMP，Internet Control Message Protocol，互联网控制消息协议。数据传输过程中出现的问题通过ICMP来反馈，如不能被递交的数据报、数据缓冲区溢出等。ICMP反馈在网络中被封装在正常的IP数据报中，发送者可以是主机或路由器。ICMP像IP一样有版本4和版本6。

（3）传输层协议

1）TCP，Transmission Control Protocol，传输控制协议。面向连接的传输协议，在两个应用进程（用所谓的端口（port）来区分）之间提供安全的通信。

2）UDP，User Datagram Protocol，用户数据报协议。面向无连接的传输协议。

3）其他协议。

4）MIME，Multipurpose Internet Mail Extension，多用途互联网邮件扩展。MIME扩展了SMTP的功能，使二进制和8Bit ASCII码文件传输成为可能。

5）DNS，Domain Name System，域名系统。邮箱、网页及其他的网络资源都有一个二进制地址。使用DNS可以用人们更容易表达的名字来标识这些资源，如http：//www.nasa.gov。可以使用一个UDP分组将这些名字发向DNS服务器，DNS服务器将这些名字翻译成二进制的地址，并将它返回给提出请求者。

6）HTTP，Hyper Transfer Protocol，超文本传输协议。用它来给互联网中World Wide Web（WWW）的资源编址定位。HTTP使同一个资源到不同的请求者的多媒体通信成为可能。

7）SNMP，Simple Network Management Protocol，简单网络管理协议。使用该协议来管理和控制一个网络，参见6.7节。

6.4.2 ISDN与B-ISDN

ISDN不是网络，而是一种业务。这种业务由基于电路交换的公共电话网提供。在瑞士，大约从1970年开始，传输设备已经开始数字化，1985年交换机数字化。但从电话设备到与之相连的交换机的本地环路还是模拟的。ISDN在这最后一段也使用数字技术。话音传输要求传统电话网的信道带宽是从300～3400Hz。由于在后续过程中，还需要在信令信道进行复用，因此在本地交换机当中就将信令分离出来。对数字化的ISDN连接来说，不需要进行这样的信令分离，因此本地环路有足够的带宽。ISDN连接（基本连接）的传输速率是144kBit/s，它被划分成2个B信道（每个B信道为64kBit/s，

用于电话、传真、数据传输等）和一个 D 信道（16kBit/s，用于控制功能）。两个 B 信道是独立的，可以同时用于两个相互独立的连接，也可以结合起来用于一个 128kBit/s 的连接。

对于大用户来说，2MBit/s 的连接也是可能的（基带连接）。

ISDN 还提供一些补充业务，如主叫号码提供等。

对多媒体应用来说 ISDN 还是太弱，因此应该使用 B-ISDN。B-ISDN 的原理同 ISDN 完全相同，但数据传输速率可达 150MBit/s。B-ISDN 能满足完全不同的业务的不同参数要求，如瞬时数据传输速率、时延等。因此 B-ISDN 所基于的传输网络要保证上述的各参数都能有效地工作，从目前来看，最好还是使用 ATM。

6.5 无线网络

6.5.1 企业无线网

企业无线网是无线站点（内嵌在交通工具）或手持对讲机设备使用的、用于进行话音通信的网络。这种网络技术已经使用了数十年，采用 VHF 和 UFH 频段的 FM（频率调制）技术。通过采用中继技术，它的覆盖范围可以扩大。

多个组织可以使用相同的网络。接入控制由用户自己通过监听信道是否被占用来完成。通过使用选择呼叫系统（NF-Tone，由接收者使用简单的开关来评估）可以提高使用的舒适度。

企业无线网最主要的缺点是为了一个空闲信道，使用者必须等较长的时间。无线频谱没有得到很好的利用，因为有时虽然有空闲的频谱，但无线终端设备却不能占用。蜂窝网络和政府机关网络在这点上明显地得到了改善。

6.5.2 蜂窝无线网

当今几乎所有人对移动通信都有很强烈的需求。在这样大的市场氛围下，只有经济、合理地使用频谱的技术才能得到推广。蜂窝网络得到普及应用得益于这样的想法：手机使用者只要能进行无线通话就行，而不管无线点发射的距离是多少。根据地形和人口密度，整个陆地被分成直径为几百米到几千米的蜂窝。每个蜂窝都有一个基站，天线放在例如一个较高的房子上。天线的位置不能放在山上，因为要尽量避免离目的地太远，以免发射距离过大。无线传输仅发生在一个蜂窝内部，之后就连接到固定网络（同轴电缆、LWL 等，部分地区可以使用定向无线连接）。由于蜂窝较小，20mW ~ 2W 这么小的发射功率就可以。通过基站，手机的发射功率可以降到最低值，以延长手机电池的使用寿命。在距离较远的蜂窝，相同的频率可以重复使用，因此整个系统可以同时服务许多用户。一个蜂窝可以有 1 ~ 8 个载频，每个载频有 8 个信道。手机采用 TDMA 方式接入信道。

蜂窝系统需要有控制功能，以解决如下的问题：

1）漫游：系统如何得知当前手机的位置？只要手机开机，它就会与离它最近的基站进行联系。基站将手机的信息发往该手机注册所在的交换机，之后手机被周期性地询问，它的位置也就被锁定了。

2）切换：用户如何在通话进行的过程中更换蜂窝？在蜂窝的边缘连接的质量会降低。基站通过固定网络询问邻居基站，是否那里的接收效果会比较好。之后如果频率转换成功，在很短的时间内连接就会切换到另一个基站。

在欧洲及其他一些地区使用 GSM（全球移动系统，但它最初的含义是指移动通信特别工作组，即标准化联盟中的工作组）系统。话音传输是数字化的，无线传输部分信号进行加密，但固定网络部分没有加密。为了节省带宽，话音信号采用信源编码，采用 LPC 编码（参见 4.1.3 节）可以将数据传输速率减少到 13.4kBit/s。随后的信道编码可以将速率提高到 22.8kBit/s。调制技术采用 GMFSK（Gaussian Minimum Frequency Shift Keying，参见 3.4.3 节）。工作频率在 UHF 频段（900MHz），这样可以使用较小的手持天线。

因为 900MHz 频段资源已经变得很紧张，因此在 1.8GHz 频段也开发了同样的系统（采用 DCS-1800 标准）。

在飞机上不允许使用手机。首先它影响飞机上的电子设备，其次它干扰座位上的邻居，第三因为它可以同时接入到多个基站，可能将蜂窝系统搞瘫痪。

值得注意的是系统的命名：在瑞士系统被称 NATEL（国家自动电话），这是 20 世纪 70 年代的表达方式。那时的设备重达几公斤，耗电量很大，因此只适合于车载。那时没人会想到在这么短的时间内设备会变得这么小、这么便宜，会这么大量地被投入使用。在德国手机被称为砖头，而在美国蜂窝网络技术被称为教父。不能设想多数用户根据这个名字会真正想到什么。

移动电话是由公司运营的无线网络的一个组成部分。相反，无绳电话是无线点到点连接的一部分，完全属于用户域。无绳电话的目的是在 100～300m 的范围内让用户自由地移动。无绳电话属于市场上的第 3 代，完全数字化并且对话音进行了加密，它使用的标准是 DECT（Digital Enhanced Cordless Telephone）。

进一步发展的产品将 DECT 与 GSM 功能结合起来。长远的目标当然是有这样的设备，首先可以选择 DECT 连接，然后寻找 GSM900MHz 或 GSM1.8GHz 网络，最后再找卫星网络。

目前已经有了第 3 代移动通信产品。UMTS（Universal Mobile Telecommunication System）包括一系列标准，是为数据通信而不是话音通信而优化的。使用 UMTS 全球范围内的多媒体通信将成为可能。电信公司为第 3 代产品执照的投资已达天文数字，此外基础设施的建设还需大规模的投资。

无线通信使用的频谱是 2G（很遗憾全球范围内并不是统一的），信道的接入控制有多种方案，如 CDMA。手机可以使用软切换同时与多个基站保持连接，而选择其中的一个进行通信。

除了网络技术之外，业务的内容和设备的手持性（可操作、显示、重量、自动运行）同价格一样对商业上的成功非常关键。

6.5.3 集群无线网

集群无线网络用于政府机关，如警察局、卫生局、消防队等。人们也称它为 BOS 无线（BOS 是具有安全任务的政府机关）。集群无线网也被应用在飞机场。每个终端都分配 10 个频率，按照管理、加油、安全、迎宾等来组织。当一个飞机开过来时，有些频率可能会出现过载，而另一个终端的信道可能会是空闲的。最好的情况是给一个机场提供 40 个频率，这些频率被动态地分配和管理。集群无线网络就是这样做的。

每个无线设备等待管理信道，这是由一个交换机来完成的。如果某人需要讲话，那么他不是占用一个特定的频率，而是向设备中输入一个地址。这个地址被传到交换机中，交换机按地址呼叫另一个设备，这样两个无线设备就可以使用一个空闲的频率进行通信。这是全自动实现的，几乎感觉不到时延。

这种功能当然使用蜂窝网络也能完成。但公共蜂窝网络在出现较大的事件时很快就会过载。虽然 GSM 网络可以给特定用户一些优先级，但集群网络可以提供以下业务：

1）每个组织单元都是一个封闭的组，尽管所有人都使用相同的物理网络。

2）跨不同组织单元之间的连接由指定的人来完成，进入到公共电话网中也是一样。

3）有组地址，因此可以实现会议通话。

4）终端设备之间可以直接通信，而无需使用基站。

5）需要时可以实现端到端的加密。

6）组织机构的集群无线设备可以执行特殊任务，基于安全上的考虑，当无线设备丢失或被偷时，可以通过遥控将它锁死。

对于集群无线网络有两种相互竞争的系统。德国和其他一些国家使用 TETRA（Trans European Trunked Radio）系统。它的容量是 100 ~ 100000 用户，用户密度为每平方千米 70 用户。净数据传输速率为 19.2kBit/s，信道接入使用 TDMA，使用 DQPSK 来调制。

在法国和瑞士使用 TETRAPOL 系统，它基于 FDMA。遗憾的是没能成功做到使用一个统一的标准。

6.5.4 卫星网络

关于卫星网络，在 5.5.7 节中已经讨论过。

6.6 数据网的测量技术

数据网的构造是很复杂的，相应地需要测量技术。正如数字技术将阴极射线示波器向逻辑分析仪推进了一步，数据网对协议分析仪有很大的影响。

数据网实际上是不同协议、不同的终端设备、不同的传输媒体、不同的节点、不同的负载状况以及不同地区应用的一个组合，因此出现不同的差错可能并不奇怪。下

面是必须要进行处理的难点和重点差错情形：协议差错（Verletzungen）、流量过载、安装和配置差错（错误的配置参数）、运行中出现损坏（瘫痪）。

测量技术最初包括能测量下面内容的设备：电平、频率响应、阻抗、线性和非线性失真、噪声、抖动、比特差错率、振荡器的相位噪声。

随着技术的发展，出现了专用的测量设备，如：接口测试仪、协议分析仪、流量模拟器、网络分析仪（可以测量数据吞吐量、时延和响应时间、分组差错率、缓冲区溢出状态等）。

协议分析仪同样也是分层结构的，可以有选择地对不同的网络层次的开销进行评估。这些设备通常有很大的内存空间，可以对对上下文环境敏感的重点内容进行存储（通过设置不同的记录文件）。

网络分析仪也提供统计输出，可以统计中、长期的数据，因此也可以作为网络规划的基础。但必须牢记的是，协议分析仪和网络分析仪提供的只是统计数据，具体来说，测量技术要能够做到：必须有足够多的探测量、探测必须具有代表性。

除此之外还应注意的是，复杂系统会出现很强的非线性特征。如果改变一个参数，例如负载增加，那么首先吞吐量就会下降，然后整个系统可能会突然崩溃。因此根据测量数据进行推断的方法是很危险的。

6.7 网络管理

网络管理完成如下两个任务：

1）确保当前网络的功能（发现损坏的元件并且更换）

2）尽量发挥网络中有效的（经济的）功能（给出网络的负载，并发现网络中可能的弱点，这样可以相应地对网络进行重新配置，或是扩容）

有效性的提高在过去的10年中已经将网络传输的费用降低了40%，而有效性的提高是通过灵活和更好地设置网络的负载来实现的。灵活性是通过不断增加的软件占有比例来实现的，这个比例从接近零（PDH）到超出50%（SDH/ATM）。

有效性可以通过下面内容产生影响：

1）由于较多的网络流量到达而产生的临时过载。可以在多个网络层次上增加协议实现流量控制来减少这类过载。因此，通过网络管理可以找出出现问题的地方，并进行处理。

2）由于网络结构产生的过载。这类过载可以通过精确的网络规划，网络监测以及网络的扩容来避免。

网络管理通过测量技术获得信息，并据此采取一定的措施。它的时间跨度可以从秒（差错定位）到年（网络扩容）。相应地，用于网络管理的投资也有不同的规模。

网络管理与现在它在网络中所能发挥的性能相比还有很大的差距。对于网络管理来说，标准化也是很重要的一项工作，这项工作正在进行。由于网络管理是关于网络的，所以针对NMT再开发一个分层模型是很正常的，如图6-30所示。

这里对每个功能进行了分层，层次的内容随网络的不同而变化。最底层只涉及网络单元（节点、传输设置等），在这之上对各个网络单元进行控制和协调，并且收集测量数据。网络管理层负责所有网络单元的管理，包括分组功能，要尽可能做到与设备商无关。例如，在对配置进行管理时，要将租用线路关掉。业务管理层负责监控是否每个业务都可用，并且负责管理用户数据。商业管理涉及企业经济上的运营活动等，并负责描绘长期目标。总的来说，低层涉及网络技术相关的内容，很具体；而高层涉及商业管理，相对较抽象。

范围包括：差错处理（差错识别、差错诊断、差错处理）、性能管理（中、长期系统的评估）、配置（过载分担、网络状态信息）、安全（有意或无意的、未经授权对管理和被管理网络的进攻）、计费。

网络管理的人员和财政花费在网络规划时是无论如何都要在考虑范围内的。

上面所介绍的 NMT 模型是 ITU 推荐的一个模型。在 TCP/IP 领域 SNMP（Simple Network Management Protocol）协议实际上已经成为一个事实上的标准。

6.8 对进一步学习的建议

对信息传输技术的介绍到这里就结束了。其中的重点不是材料的深度，而是它的广度，这是因为通信领域多方面之间的相互关系要求基础牢固的全能型人才。本书的目的有 3 个：

1）传播整个领域上的必要的基础知识；

2）培养自学的能力；

3）增强对电信技术的爱好和兴趣。

这 3 点对保证今后继续学习较深入的内容以及毕业后的若干年内不断补充本领域的最新内容是很必要的。继续学习可能还需要先行对一些数学基础知识进行深化，特别值得一提的是：

1）对信息论和调制类型中的干扰特性而言，还需要概率论和统计学的相关知识；

2）对交换技术和通信网而言，需要排队论和马尔可夫过程的相关知识；

3）对信道编码和加密而言，需要线性代数和伽罗瓦域算法的相关知识；

4）对波理论和高频技术而言，需要偏微分方程和矢量分析的相关知识；

5）对描述和实现复杂的系统单元，例如可适配均衡器，需要系统论的相关知识。

下面是一些对可能的继续学习和入门的建议。

1）对于全面的学习我推荐一些概括介绍和指导类型的书籍，如参考文献【Mei92】、【Loc97】和【Her00】。参考文献【Pro94】和【Skl98】与上面所列书籍内容有些重复，但使用专业英语，对上述内容是一个很好的深化。

2）点到点传输还必需包括对噪声的分析，为此应该推荐参考文献【Con86】、【Mil97】、【Lük92】、【Kam92】、【Kro96】和【Klo01】。

3）关于通信网的更多知识主要可以参考参考文献【Pet00】、【Hal96】、【Sta97】、【Tan97】和【Ger94】。

4）信息论和编码方面可以深入学习参考文献【Mil90】、【Swe92】和【Bos92】。

5）参考文献【Mei92】和【Ger97】包括许多对这里所介绍概念的实现的建议，这需要一些电子电路、数字技术和信号处理方面的知识。

6）如果想要了解关于计算机方面的内容，参考文献【Kam01】和【Ger97】是很好的选择。

本书符号释义表

符号	含义
$a(t)$	包络
a_F	以 dB 为单位的自由空间衰减系数
A	幅度
A_D	自由空间衰减
B	带宽
B_a	模拟信号的带宽
B_N	奈奎斯特带宽
B_u	有效传输带宽
C	以 bit/s 为单位的信道能力
d	码间距
D	动态性
D	地理上的连接距离
e	一个码字中的误码数
$e(t)$	误差信号(error)
E	电场强度
E_{bit}	每 bit 的能量
$E(s)$	期望值
f	频率
F	噪声系数
F_{ant}	天线噪声系数
$g(t)$	单位响应
G	增益
G	发生器矩阵
G_M	以 dB 为单位的调制增益
G_C	以 dB 为单位的编码增益
h	码的汉明距离
h	地形上的高度差
$h(t)$	脉冲响应
H	信源熵
H	监测矩阵
H	磁场强度
H_0	判决内容
$H(j\omega)$	频率响应，$h(t)$的傅里叶变换
$H(s)$	传输函数，$h(t)$的拉普拉斯变换
I	用 bit 表示的信息量
J	用 bit/s 表示的信息速率
J_n	第一类 n 阶贝塞尔函数
k	ADC 或 DAC 的字宽
k	一个码字的校验位的数目
k	玻尔兹曼常数
K	常量
L	电平(Level)
l	长度
m	调制深度(AM)
m	一个码字中的有用位的数量
M	码或调制的进制数
n	一个码字的有用位或校验位的数量
n	噪声系数
N_0	热噪声功率密度
P_S	信号功率
P_N	干扰功率，噪声功率
$p(s)$	s 的概率密度函数
p_{Fehler}	误差概率
p_{Bit}	误码概率
p_{Block}	组(块)误差概率
p_R	残留误差概率
q	量化间隔
R	用 Bit/s 表示的数据传输速率
R	冗余
R	地球半径

（续）

符号	含　　义	符号	含　　义
R	苏黎士相对数	w	码字的权重
R_C	编码率	$x(t)$	信号，输入信号
R_s	平滑的苏黎士相对数	$X(\omega)$	$x(t)$的傅里叶谱
$s(t)$	信号	$X(s)$	$x(t)$的拉普拉斯谱
$s_k(t)$	信号的同相成分	$y(t)$	信号，输出信号
$s_{Na}(t)$	信息信号	$Y(\omega)$	$y(t)$的傅里叶谱
$s_m(t)$	已调信号	$Y(s)$	$y(t)$的拉普拉斯谱
$s_q(t)$	信号的正交成分	Y	亮度信号（TV）
S	符号率，波特率	Z_W	波阻抗
S	坡印廷（Poynting）矢量	β	带宽放大系数
SF	太阳通量	$\delta(t)$	迪拉克（Dirac）冲击
SR_A	以 dB 为单位的解调器/解码器输出端的信噪比	$\varepsilon(t)$	阶跃函数，阶梯函数
SR_K	以 dB 为单位的信道末端（解调器输入端）的信噪比	ε_0	真空的介电常数
SR_Q	以 dB 为单位的量化信噪比	λ	波长
$s_{Tr}(t)$	载波信号	μ	调制指数（FM）
T	时间	μ_0	真空的绝对透气性（渗透性，磁导率）
T	周期	τ	时延
T_{Bit}	一个 Bit 的持续时间	ϕ_{Tr}	载波-零相位
T_R	噪声温度	$\Psi(t)$	三角函数的参数
u	电压	ω	角频率
U	色差信号（TV）	ω_{Tr}	载波的角频率
V	色差信号（TV）	○—○	傅里叶、拉普拉斯或 z 对应项

缩 略 语

缩写	全 拼	释 义
AAL	ATM Adaption Layer	ATM 适配层
AC	Alternate Current, Wechsel(strom)signal	交流电,交变信号
ACELP	Algebraic Code Excited Linear Prediction	代数码激励线性预测
ADC	Analog-Digital-Wandlung, Analog-Digital-Wandler	模/数转换,模/数转换器
ADM	Adaptive Delta Modulation	自适应增量调制
ADM	Add Drop Multiplexer	分插复用器
ADSL	Asymmetrical Digital Subscriber Line	非对称数字用户线
AES	Advanced Encryption Standard	高级加密标准
AF	Audio Frequencies, Nieder Frequenz (NF)	音频频率,低频
AFSK	Audio Frequency Shift Keying	音频频移键控
AGC	Automatic Gain Control	自动增益控制
AIS	Alarm Indication Signal	告警指示信号
AKF	Auto Korrelations Funktion	自动交织函数
ALE	Automatic Link Establishment	自动链路建立
AM	Amplitude Nmodulation	幅度调制
AMI	Alternate Mark Inversion	交替信号反转
AMTOR	Amateur Teletype Over Radio	业余无线电爱好者无线电传打字机
ANSI	American National Standards Institute	美国国家标准化协会
AP	All Pass	全通
APD	Avalanche Photo Diode	雪崩光敏二级管
APSK	Amplitude Phase Shift Keying	幅度相移键控
ARI	Autofahrer Rundfunk Information	针对机动车司机的无线广播信息
ARQ	Automatic Repeat Request	自动重传请求
ASCII	American Standard Code for Information Interchange	美国信息互换标准代码
ASK	Amplitude Shift Keying, Amplitudenumtastung	幅移键控
ATM	Asynchronous Transfer Mode	异步传输模式
ATSC	Advanced Television Systems Committee	高级电视业务顾问委员会
AWGN	Additive White Gaussian Noise	加性高白斯噪声
AZ	Anschluss Zentrale	接入交换机
BAS	Bild, Austastlticke, Synchronisation	图像、消隐间隔同步

（续）

缩写	全 拼	释 义
BB	Basis Band	基带
BER	Bit Error Ratio, Bitfehlerquote	误码率
BFO	Beat Frequency Oscillator	拍频振荡器
BGP	Border Gateway Protocol	边界网关协议
BITE	Built In Test Equipment	内置测试设备
BOS	Behörden und Organisationen mit Sicherheitsaufgaben	具有安全任务的政府机关
BP	Band Pass	带通
CAP	Carrierless AM/PM	无载频 AM/PM
CATV	Cable Television, Community Antenna Television	有线电视，社区天线电视
CCITT	Comité Consultatif International Télégraphique et Téléphonique	国际电报电话咨询委员会
CCIR	Comité Consultatif International des Radiocommunications	国际无线通信咨询委员会
CD	Compact Disc	紧密光盘
CDM	Code Division Multiplex	码分复用
CDMA	Code Division Multiple Access	码分多址接入
CEPT	Conférence Ettropéen des Administrations des Postes et des Télécommunications	欧洲邮电管理大会
CIF	Common Interleaved Frame	公共交织帧
CMI	Coded Mark Inversion	信号反转码
COFDM	Coded Orthogonal Frequency Division Multiplex	编码正交频分复用
CPFSK	Continous Phase Frequency Shift Keying	连续相位频移键控
CRC	Cyclic Redundancy Check	循环冗余检验
CSMA	Carrier Sense Multiple Access	载波监听多址接入
CSMA/CD	Carrier Sense Multiple Access with Collision Detection	带有冲突检测的载波监听多址接入
CS-ACELP	Conjugate-Structure Algebraic Code Excited Linear Prediction	共轭代数编码激励线性预测
CVSD	Continuously Variable Slope Delta Modulation	连续可变斜率增量调制
DAB	Digital Audio Broadcasting	数字音频广播
DAMA	Demand Assignment Multiple Access	按需多址接入
DAT	Digital Audio Tape	数字音频磁带
DC	Direct Current, Gleich(strom)signal	直流点，等(电流)信号
DCDM	Digitally Controlled Delta Modulation	数控增量调制
DCE	Data Communication Equipment, Daten-übertragungsgerät	数字通信设备，数据传输设备

（续）

缩写	全　拼	释　义
DCO	Digitally Controlled Oscillator	数控振荡器
DCT	Diskrete Cosinus Transformation	离散余弦变换
DDS	Direct Digital Synthesis	直接数字频率合成器
DECT	Digital Enhanced Cordless Telephone	数字增强无绳电话
DES	Data Encryption Standard	数据加密标准
DFT	Diskrete Fourier-Transformation	离散傅里叶变换
DM	Delta Modulation	增量调制
DMT	Discrete Multi-Tone	离散多音调
DMUX	Demultiplexer, Demultiplexierung	解复用器，解复用
DNS	Domain Name System	域名系统
DPCM	Differentielle PCM	差分 PCM
DQDB	Distributed Queue Dual Bus	分布式队列双总线
DRM	Digital Radio Mondiale	数字无线广播
DSI	Digital Speech Interpolation	数字话音插空
DSP	Digitaler Signal Prozessor	数字信号处理
DSR	Digital Satellite Radio	数字卫星广播
DSSC	Double Sideband Suppressed Carrier	双边带抑制载波
DSSS	Direct Sequence Spread Spectrum	直接序列扩频
DTE	Data Teminal Equipment, Daten-Endgerät	数据终端设备
DUT	Device Under Test (Prüfling)	被测设备
DVB	Digital Video Broadcasting	数字视频广播
DWDM	Dense WDM	密集 WDM
ECSS	Extracted Carrier Selectable Sideband	载波提取可选边带
EDFA	Erbium-Doped Fibre Amplifiers	掺铒光纤放大器
EDTV	Extended Definion Television	增强清晰度电视
EOW	Engineering Order Wire, Dienstkanal	工程联络线，业务信道
ERP	Effective Radiated Power	有效辐射功率
ESB	Ein Seiten Band	单边带
ETSI	European Telecommunication Standards Institute	欧洲电信标准化协会
FBAS	Farbe, Bild, Austastlücke, Synchronisation	颜色、图像、消隐间隔同步
FDD	Frequency Division Duplexing	频分双向连接
FDDI	Fibre Distributed Data Interface	光纤分布数字接口
FDM	Frequency Division Multiplex, Frequenz-Multiplex	频分复用
FDMA	Frequency Division Multiple Access	频分多址接入

（续）

缩写	全 拼	释 义
FEC	Forward Error Correction	前向纠错
FEXT	Far End Crosstalk, Fernübersprechen	远端串扰
FFSK	Fast Frequency Shift Keying	快速频移键控
FFT	Fast Fourier Transform	快速傅里叶变换
FHSS	Frequency Hopping Spread Spectrum	跳频扩频
FIC	Fast Information Channel	快速信息信道
FIR	Finite Impulse Response	有限脉冲响应
FITL	Fibre In The Loop	本地环路光纤
FM	Frequenz Modulation	频率调制
FOT	Frequency of Optimal Traffic	最优流量频率
FSK	Frequency Shift Keying, Frequenzumtastung	频移键控
FT	Fourier Transformation	傅里叶变换
FTP	File Transfer Protocol	文件传输协议
FTTB	Fibre To The Building	光纤到楼
FTTC	Fibre To The Curb	光纤到路
FTTH	Fibre To The Home	光纤到家
GF	Galois Feld	伽罗瓦(Galois)场
GMII	Gigabit Media Independent Interface	吉比特媒体无关接口
GMSK	Gaussian Minimum Shift Keying	高斯最小键控
GPS	Global Positionning System	全球定位系统
GSM	Global System for Mobile Communication	全球无线移动通信系统
HDB-3	High Density Bipolar Of Order 3	3 阶高密度双极性码
HDLC	High-Level Data Link Control	高级数据链路控制
HDSL	High Speed Digital Subscriber Line	高速数字用户线
HDTV	*High Definition Television*	高清晰度电视
HIDM	High Information Delta Modulation	高信息增量调制
HP	Hoch Pass	高通
HP	High Profile	明显配置文件
HTTP	Hyper Text Transfer Protocol	超文本传输协议
ICMP	Internet Control Message Protocol	互联网控制消息协议
IDEA	International Data Encryption Algorithm	国际数据加密算法
IEC	International Electrical Commission	国际电工委员会
IEEE	Institute of Electrical and Electronic Engineers	电气和电子工程师协会
IIR	Infinite Impulse Response	无限脉冲响应

（续）

缩写	全　　拼	释　　义
IF	Intermediate Frequency, Zwischenfrequenz	中间频率
IFS	Integriertes Fernmelde System	集成通信系统
IDCT	Inverse Diskrete Cosinus Transformation	反向离散余弦变换
IDFT	Inverse Diskrete Fourier Transformation	反向离散傅里叶变换
IM	Intensitäts Modulation	亮度调制
IOL	Inter-Orbit Link	轨道间链路
IP	Internet Protocol	互联网协议
ISB	Independent Side Band	独立边带
ISDN	Integrated Services Digital Network	综合业务数字网
ISI	Inter Symbol Interference, Pulsübersprechen	码间干扰，码间串扰
ISL	Inter Satellites Link	卫星间链路
ISO	International Standards Organization	国际标准化组织
ITU	International Telecommunication Union	国际电信联盟
JPEG	Joint Photographic Experts Group	联合图像专家组
KKF	Kreuzkorrelationsfunktion	互相关函数
KW	Kurz Welle	短波
LAN	Local Area Network	局域网
LAPB	Link Access Procedure, Balanced	平衡链路接入过程
LAPD	Link Access Procedure D Channel	D 信道链路接入过程
LD	Laser Diode	激光二极管
LD-CELP	Low Delay Code Excited Linear Prediction	低时延编码加速线性预测
LDTV	Low Definition Television	低清晰度电视
LED	Light Emitting Diode, Leuchtdiode	发光二极管
LEO	Low Earth Orbit	低地球轨道
LF	Loop Filter	环路滤波器
LLC	Logical Link Control	逻辑链路控制
LM	lineare Modulation	线性调制
LNA	Low Noise Amplifier	低噪声放大器
LNB	Low Noise Block	低噪声块
LNC	Low Noise Converter	低噪声转换器
LO	Lokal-Oszillator	本地振荡器
LPC	Linear Predictive Coding	线性预测编码
LRC	Longitudinal Redundancy Check	经度冗余检测
LSB	Lower Side Band, unteres Seitenband	低边带，下边带

（续）

缩写	全　拼	释　义
LSB	Least Significant Bit(niederwertigstes Bit)	最低(有效)位
LTD	Linear Discrete Timeinvariant	线性离散时不变
LTI	Linear Time Invariant	线性时不变
LUF	Lowest Usable Frequency	最低可用频率
LW	Lang Welle	长波
LWL	Licht Wellen Leiter	光波导
MAC	Media Access Control	媒体接入控制
MAN	Metropolitan Area Network	城域网
MBC	Meteor Burst Communication	流星突发通信
MELP	Mixed Excitation Linear Prediction	混合加速线性预测
MII	Media Independent Interface	媒体无关接口
MIME	Multipurpose Internet Mail Extension	多用途互联网邮件扩充协议
MP	Main Profile	主配置文件
MPEG	Moving Picture Experts Group	运动图像专家组
MR-QAM	Multi Resolution QAM	多解析 QAM
MSB	Most Significant Bit(höchstwertigstes Bit)	最高(有效)位
MSC	Main Service Channel	主业务信道
MSK	Minimum Shift Keying	最小键控
MUF	Maximum Usable Frequency	最大可用频率
MUSICAM	Masking Pattern Universal Subband Integrated Coding And Multiplexing	伪装模式通用子带集成编码和复用
MUX	Multiplexer, Multiplexierung	复用器,复用
MW	Mittel Welle	中波
NATEL	Nationales Autotelefon	国家自动电话机
NEXT	Near End Crosstalk, Nahübersprechen	近端串扰
NF	Nieder Frequenz	低频
NMT	Network Management	网络管理
NRZ	Non Return to Zero	非归零
NSA	National Security Agency	国家安全局
NTSC	National Television System Committee	国家电视系统委员会
OFDM	Orthogonal Frequency Division Multiplex	正交频分复用
OOK	On-Off-Keying	开关键控
OSB	Oberes Seiten Band	上边带
OSI	Open Systems Interconnection	开放系统互联

（续）

缩写	全　拼	释　义
OSPF	Open Shortest Path First	开放最短路径优先
OQPSK	Offset Quadrature Phase Shift Keying	偏移正交相位键控
PAD	Programme Associated Data	节目相关数据
PAL	Phase Alternating Line	逐行倒相
PAM	Puls Amplituden Modulation	脉冲幅度调制
PDH	Plesiochrone Digitale Hierarchie	准同步数字序列
PCM	Pulse Code Modulation	脉冲编码调制
PD	Phasen Detektor	相位检测器
PDM	Puls-Dauer-Modulation	脉冲时长调制
PFM	Puls-Frequenz-Modulation	脉冲频率调制
PGP	Pretty Good Privacy	相当好的私密性
PLC	Power Line Communication	电力线通信
PM	Phasen Modulation	相位调制
POTS	Plaine Old Telephone Service	简单电话业务
PPM	Puls-Phasen-Modulation, Puls-Positions-Modulation	脉冲相位调制
PRBN	Pseudo Random Binary Noise	伪随机二进制噪声
PSK	Phase Shift Keying, Phasenumtastung	相移键控
PWM	Puls-Width-Modulation, Puls-Breiten-Modulation	脉冲宽带调制
QAM	Quadrature Amplitude Modulation	正交幅度调制
QPAM	Quantisierte PAM	量化的 PAM
QPSK	Quadrature Phase Shift Keying	回相相移键控
RDS	Radio Data System	无线数据系统
RF	Radio Frequencies, Hochfrequenz(HF)	无线电频率，高频
RFM	Rechteck Frequenz Modulation	直方波频率调制
RITL	Radio In The Loop	无线本地环路
RTTY	Radio Teletype	无线电传打字机
RX	Receiver, Empfänger	接收器
RZ	Return to Zero	归零
SAP	Service Access Point	业务接入点
SB-ADPCM	Sub-Band Adaptive Differential Pulse Code Modulation	子带可适配差分脉冲编码调制
SDH	Synchrone Digitale Hierarchie	同步数字序列
SDTV	Standard Definition Television	标准清晰度电视
SFN	Single Frequency Network	单频网络
SHA	Secure Hash Algorithm	安全哈希算法

（续）

缩写	全　拼	释　义
SID	Sudden Ionospheric Disturbance	电离层的突然扰动
SITOR	Simplex Teletype Over Radio	简单无线电传打字机
SMTP	Simple Mail Transfer Protocol	简单邮件传输协议
SNMP	Simple Network Management Protocol	简单网络管理协议
SNR	Signal to Noise Ratio	信噪比
SONET	Synchronous Optical Network	同步光网络
SP	Simple Profile	简单配置文件
SSB	Single Side Band, Einseitenbandmodulation	单边带，单边带调制
SSP	Spatial Scalable Profile	空间可扩展配置文件
SSSC	Single Side Band Suppressed Carrier (identisch zu SSB)	单边带抑制载波
SS7	Signalisierungssystem No. 7	No. 7 信令系统
STM	Synchronous Transport Modul	同步传输模式
STP	Shielded Twisted Pair	屏蔽双绞线
S&H	Sample & Hold, Halteglied	采样和保持
TASI	Time Assignement Speech Interpolation	按时话音交织
TCM	Trellis Codierte Modulation	栅格编码调制
TCP	Transmission Control Protocol	传输控制协议
TCP/IP	Transmission Control Protocol/Internet Protocol	传输控制协议/互联网协议
TDD	Time Division Duplexing	时分双向通信
TDM	Time Division Multiplex, Zeitmultiplex	时分复用
TDMA	Time Division Multiple Access	时分多址接入
TETRA	Trans European Trunked Radio	跨欧洲中继无线电
TFM	Tamed Frequency Modulation	简单频率调制
TOR	Teletype Over Radio	无线电传打字机
TP	Tief Pass	低通
TX	Transmitter, Sender	发送器
TRCV	Transceiver, Sende-Empfänger	收发器
TVA	Teilnehmervermittlungsanlage, Haustelefonzentrale	用户交换机设备
UART	Universal Asynchronous Receiver Transmitter	通用异步收发器
UDP	User Datagram Protocol	用户数据报协议
UKV	Universelle Kommunikations Verkabelung	通用通信布线
UKW	Ultra Kurz Wellen	超短波

（续）

缩写	全　拼	释　义
UMTS	Universal Mobile Telecommunication System	通用无线移动通信系统
USB	Upper Side Band, unteres Seitenband（Bedeutung je nach Sprache!）	上边带
UTP	Unshielded Twisted Pair	非屏蔽双绞线
VCO	Voltage Controlled Oscillator	压控振荡器
VFC	Voltage to Frequency Converter	电压频率转换器
VHDSL	Very High Bitrate Digital Subscriber Line	非常高速数字用户线
VRC	Vertical Redundancy Check	垂直冗余检测
VSB	Vestigual Side Band, Restseitenbandmodulation	残边带
VZAUA	Verein Zur Abschaffung Unnötiger Abkürzungen	消除不必要所需联盟
WAN	Wide Area Network	广域网
WRC	World Radio Conference	世界无线电大会
WDM	Wavelength Division Multiplex	波分复用
WM	Winkel Modulation	相角调制
WWW	World Wide Web	万维网
XPIC	Cross Polarization Interference Canceller	交叉极化干扰消除器
ZF	Zwischen Frequenz	中间频率

参考文献

[Abm94] Abmayr, W.: Einführung in die digitale Bildverarbeitung
Teubner-Verlag, Stuttgart, 1994

[Bän02] Bäni, W.: Wavelets – Eine Einführung für Ingenieure
Oldenbourg-Verlag, München, 2002

[Bes93] Best, R.: Theorie und Anwendungen des Phase-locked-Loops
AT-Verlag, Aarau, 1993

[Beu94] Beutelspacher, A.: Kryptologie
Vieweg-Verlag, Braunschweig/Wiesbaden, 1994

[Bos92] Bossert, M.: Kanalcodierung
Teubner-Verlag, Stuttgart, 1992

[Con86] Connor, F.R.: Rauschen
Vieweg-Verlag, Braunschweig/Wiesbaden, 1986

[Gen98] Geng, N., Wiesbeck, W.: Planungsmethoden für die Mobilkommunikation
Springer-Verlag, Berlin, 1998

[Ger94] Gerdsen, P., Kröger, P.: Kommunikationssysteme 1
Springer-Verlag, Berlin, 1994

[Ger97] Gerdsen, P., Kröger, P.: Digitale Signalverarbeitung in der Nachrichtenüber-*tragung*. Springer-Verlag, Berlin, 1997

[Gon93] Gonzales, R.C., Woods, R.E.: Digital Image Processing
Addison-Wesley, Reading, 1993

[Hal96] Halsall, F.: Data Communications, Computer Networks and Open Systems
Addison-Wesley, Harlow, 1996

[Her00] Herter, E., Lörcher, W.: Nachrichtentechnik
Hanser-Verlag, München, 2000

[Hof97] Hoffmann, M.: Hochfrequenztechnik
Springer-Verlag, Berlin, 1997

[Höl86] Hölzler, E., Holzwarth, H.: Pulstechnik, Band 1
Springer-Verlag, Berlin, 1986

[Joh92] Johann, J.: Modulationsverfahren
Springer-Verlag, Berlin, 1992

[Jon02] Jondral, F., Machauer, R., Wiesler, A.: Software Radio
J. Schlembach Fachverlag, Weil der Stadt, 2002

[Kam92] Kammeyer, K.D.: Nachrichtenübertragung
Teubner-Verlag, Stuttgart, 1992

[Kam01] Kammeyer, K.D., Kühn, V.: MATLAB in der Nachrichtentechnik
J. Schlembach Fachverlag, Weil der Stadt, 2001

[Kon94] Kondoz, A.M.: Digital Speech (Coding for Low Bitrate Communication Systems)
John Wiley & Sons, Chichester, 1994

[Klo01] Klostermeyer, R.: Digitale Modulation
Vieweg-Verlag, Braunschweig/Wiesbaden, 2001

[Kro96] Kroschel, K.: Statistische Nachrichtentheorie
Springer-Verlag, Berlin, 1996

[Loc97] Lochmann, D.: Digitale Nachrichtentechnik
Verlag Technik, Berlin, 1997

[Lük92] Lüke, H.D.: Signalübertragung
Springer-Verlag, Berlin, 1992

[Mäu91] Mäusl, R.: Digitale Modulationsverfahren
Hüthig-Verlag, Heidelberg, 1991

[Mäu91a] Mäusl, R., Schlagheck E.: Messverfahren in der Nachrichten - Übertragungstechnik
Hüthig-Verlag, Heidelberg, 1991

[Mäu91b] Mäusl, R.: Fernsehtechnik
Hüthig-Verlag, Heidelberg, 1991

[Mäu92] Mäusl, R.: Analoge Modulationsverfahren
Hüthig-Verlag, Heidelberg, 1992

[Mei92] Meinke, Gundlach: Taschenbuch der Hochfrequenztechnik
Hrsg. von Lange, K., Löcherer, K.-H., Springer-Verlag, Berlin, 1992

[Mey00] Meyer, M.: Signalverarbeitung
Vieweg-Verlag, Braunschweig/Wiesbaden, 2000

[Mey02] Meyer, M.: Grundlagen der Informationstechnik
Vieweg-Verlag, Braunschweig/Wiesbaden, 2002

[Mil90] Mildenberger, O.: Informationstheorie und Codierung
Vieweg-Verlag, Braunschweig/Wiesbaden, 1990

[Mil97] Mildenberger, O.: Übertragungstechnik
Vieweg-Verlag, Braunschweig/Wiesbaden, 1997

[Nee00] van Nee, R., Ramjee, P.: OFDM For Wireless Multimedia Communications
Artech House, Boston/London, 2000

[Pet00] Peterson, L.L., Davie, B.S.: Computernetze
dpunkt.verlag, Heidelberg, 2000

[Pro94] Proakis, J.G., Salehi, M.: Communication Systems Engineering
Prentice-Hall, Englewood Cliffs, New Jersey, 1994

[Rei97] Reimers, U.: Digitale Fernsehtechnik
Springer-Verlag, Berlin, 1997

[Sch96] Schneider, B.: Applied Cryptography
John Wiley & Sons, New York, 1996

[Sin00] Singh, S.: Geheime Botschaften
Hanser-Verlag, München, 2000

[Skl88] Sklar, B.: Digital Communications, Fundamentals And Applications
Prentice-Hall, Englewood Cliffs, New Jersey, 1988

[Sta97] Stallings, W.: Data and Computer Communications
Prentice Hall, Upper Saddle River, New Jersey, 1997

[Stä98] Stähli, F.: Ingenieurethik an Fachhochschulen
Fortis Verlag FH / Sauerländer, Aarau, 1998

[Ste82] Steinbuch, K., Rupprecht, W.: Nachrichtentechnik, Band 2
Springer-Verlag, Berlin, 1982

[Str00] Strutz, T.: Bilddatenkompression
Vieweg-Verlag, Braunschweig/Wiesbaden, 2000

[Swe92] Sweeney, P.: Codierung zur Fehlererkennung und Fehlerkorrektur
Hanser-Verlag, München, 1992

[Tan97] Tanenbaum, A.S.: Computernetzwerke
Prentice Hall, München, 1997

[Tra93] Trachsel, R.: Ein halbes Jahrhundert Telekommunikation in der Schweiz
Sauerländer-Verlag, Aarau, 1993

[Wol97] Wolff, I.: Maxwellsche Theorie
Springer-Verlag, Berlin, 1997

[Zin90] Zinke, O., Brunswig u.a.: Hochfrequenztechnik, Band 1
Springer-Verlag, Berlin, 1990

[Zin93] Zinke, O., Brunswig u.a.: Hochfrequenztechnik, Band 2
Springer-Verlag, Berlin, 1993

国际视野　科技前沿

国际信息工程先进技术译丛

《现代通信原理》（原书第2版）
《认知无线电网络》
《高速数字系统的信号完整性和辐射发射》
《UMTS中的LTE：基于OFDMA和SC-FDMA的无线接入》
《生物医学工程学概论》（原书第2版）
《全面的功能验证：完整的工业流程》
《无线Mesh网络架构与协议》
《UMTS蜂窝系统的QoS与QoE管理》
《半导体制造与过程控制基础》
《WCDWA原理与开发设计》
《下一代移动系统：3G/B3G》
《IMS:IP多媒体概念和服务》（原书第2版）
《下一代无线系统与网络》
《深入浅出UMTS无线网络建模、规划与自动优化：理论与实践》
《HSDPA/HSUPA技术与系统设计——第三代移动通信系统宽带无线接入》
《无线传感器及元器件：网络、设计与应用》
《印制电路板——设计、制造、装配与测试》
《IPTV与网络视频：拓展广播电视的应用范围》
《多电压CMOS电路设计》
《微电子技术原理、设计与应用》
《蜂窝网络高级规划与优化2G/2.5G/3G/…向4G的演进》
《基于蜂窝系统的IMS——融合电信领域的VoIP演进》
《无线网络中的合作原理与应用》
《电生理学方法与仪器入门》
《移动电视：DVB-H、DMB、3G系统和富媒体应用》
《环境网络：支持下一代无线业务的多域协同网络》
《基于射频工程的UMTS空中接口设计与网络运行》
《未来UMTS的体系结构与业务平台：全IP的3G CDMA网络》
《UMTS-HSDPA系统的TCP性能》
《宽带无线通信中的空时编码》
《数字图像处理》（原书第4版）
《基于4G系统的移动服务技术》
《吉规模集成电路互连工艺及设计》
《高性能微处理器电路设计》

上架指导：工业技术/通信工程



ISBN 978-7-111-32505-5
封面设计：马精明
定价：88.00元

地址：北京市百万庄大街22号
邮政编码：100037
电话服务
网络服务
社服务中心：(010)88361066
门户网：http://www.cmpbook.com
销售一部：(010)68326294
教材网：http://www.cmpedu.com
销售二部：(010)88379649
读者服务部：(010)68993821
封面无防伪标均为盗版